2003

# Carbonic anhydrase and carbonic anhydrase like genes of Chlamydomonas reinhardtii

Mautusi Mitra
*Louisiana State University and Agricultural and Mechanical College*, mmitra@lsu.edu

# CARBONIC ANHYDRASE AND CARBONIC ANHYDRASE LIKE GENES OF *CHLAMYDOMONAS REINHARDTII*

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Mautusi Mitra
B.Sc., University of Calcutta, 1989
M.Sc., University of Calcutta, 1991
December, 2003

# ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to Dr. J.V. Moroney, my major advisor, for his guidance, support and patience during my course of study. I would like to thank the members of my graduate committee, Drs. Sue. G. Bartlett, David J. Longstreth, Thomas S. Moore and my outside committee member, Dr. Don Labonte for their advice regarding the project and the dissertation. I gratefully acknowledge the cooperation of my colleagues, both past and present members of Dr. Moroney's laboratory: Dr. Steve V. Pollock, Dr. Sergio Colombo, Catherine B. Mason, Ruby A.Ynalvez and Scott M. Lato. I would also like to thank Ying Xiang (Department of Biological Sciences, Louisiana State University) for the electron microscopy work for the immunolocalization of Cah6 and Dr. Göran Samuelsson (Umeå University, Sweden) for providing me with the p*MalE-Cah3* construct and the Cah3 and β-mitochondrial CA antibodies.

I am indebted to all my friends and well wishers who have tolerated my eccentricities for all these years and have never abandoned me. Finally, I would like to thank my parents for their encouragement and continuous support and above all, their unconditional love. Without them, this document would never have been completed.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ABC | ATP binding cassette |
| ATP | Adenosine triphosphate |
| Amp | Ampicillin |
| AZ | Acetazolamide |
| BBY | Berthold-Babcock-Yocum PSII particles |
| BSA | Bovine serum albumin |
| CA | Carbonic anhydrase |
| CAs | Carbonic anhydrases |
| CAM | Crassulacean acid metabolism |
| CAP | Contig assembly program |
| $C_i$ | Inorganic carbon |
| CCM | $CO_2$ concentrating mechanism |
| cDNA | Complementary DNA |
| Cys | Cysteine |
| dCTP | Deoxycytosine triphosphate |
| dNTP | Mixture of dATP, dCTP, dGTP and dTTP |
| DEPC | Diethyl pyrocarbonate |
| DTNB | 5´, 5´-dithiobis(2-nitrobenzoic azide) |
| DTT | Dithiothreitol |
| EDTA | ethylenediaminetetraacetic acid |
| EZ | Ethoxyzolamide |
| EPPS | 4-(2-hydroxyethyl)-1-piperazine propane sulfonic acid |
| EST | Expressed sequence tag |

| | |
|---|---|
| GAP | Glyceraldehyde-3-phosphate |
| Gclp | Gamma CA like protein |
| HCR | High $CO_2$ requiring phenotype |
| IPTG | Isopropyl-β, D-thiogalactopyranoside |
| kDa | Kilo dalton |
| LB | Luria-Bertani |
| Mer | 2- Mercaptoethanol |
| MIN | Minimal media |
| MS | Murashige-Skoog |
| MOPS | Morpholinepropanesulfonic acid |
| $NAD^+$ | β- nicotinamide adenine dinucleotide, oxidized form |
| $NADP^+$ | β- nicotinamide adenine dinucleotide phosphate, oxidized form |
| NADPH | β- nicotinamide adenine dinucleotide phosphate, reduced form |
| ORF | Open reading frame |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate saline buffer |
| PCR | Polymerase chain reaction |
| 3-PGA | 3-phosphogylcerate |
| ppm | Parts per million |
| PS | Photosystem |
| PSR | Proton shuttle residue |
| RACE | Rapid amplification of cDNA ends |
| RNAi | RNA interference |
| rbcL | Rubisco large subunit |

| | |
|---|---|
| Rubisco | Ribulose 1, 5 bisphosphate carboxylase /oxygenase |
| RuBP | Ribulose 1, 5 bisphospate |
| SDS | Sodium dodecyl sulfate |
| TAP | Tris Acetate Phosphate |
| WAU | Wilbur and Anderson unit |
| WT | Wild type |
| X-Gal | 5-bromo-4-chloro-3-indolyl-β-D-galactoside |

# ABSTRACT

Carbonic anhydrase (CA) is a zinc containing metalloenzyme that catalyzes the reversible interconversion of $CO_2$ and $HCO_3^-$. There are three evolutionarily unrelated CA families designated α-, β-, and γ-CA. Vertebrates have members of the α-CA family, while higher plants, algae and cyanobacteria have members belonging to all three CA families. In the green alga, *Chlamydomonas reinhardtii*, five CAs have previously been identified including three α-CAs and two β-CAs. This dissertation describes the identification and characterization of new CA genes from *C. reinhardtii*. Four new CA or CA-like genes have been discovered including two β-CAs and two γ-CAs. Three CAs were investigated further including the α-CA Cah3, one of the new β-CAs, Cah6; and a new γ-CA designated Gclp1 for gamma-CA-like protein.

Cah3 is an α-CA located in the thylakoid. Past studies with two Cah3 mutants, *ca-1* and *cia3* have shown that Cah3 plays an important role in the $CO_2$ concentrating mechanism. In this work, the mature Cah3 protein was overexpressed as a fusion protein in *E. coli* and found to have significant CA activity. This is the first report of detection of CA activity in the Cah3 protein and its partial biochemical characterization.

A novel β-CA (*Cah6)* and a putative γ-CA (*Gclp1*) gene were identified in *C. reinhardtii*. Gclp1 is one of two putative γ-CAs found in *C. reinhardtii*. Both the Cah6 and Gclp1 open reading frames (ORFs) were cloned in the overexpression vector pMal-c2x and expressed as recombinant fusion proteins. The purified Cah6 had significant *in vitro* CA activity but Gclp1 did not. Gclp1 was designated as a γ-CA like protein because it lacked detectable CA activity. Cah6 has a leader sequence consistent with a chloroplast localization. Although Cah6 is constitutively expressed under low and high $CO_2$ conditions, it is slightly upregulated under low $CO_2$ conditions. Immunolocalization studies confirmed that the Cah6 is localized to the

chloroplast stroma particularly in the starch sheath around the pyrenoid. A possible role of Cah6 in the $CO_2$ concentrating mechanism and photosynthesis is discussed.

# CHAPTER 1

# INTRODUCTION

## CARBONIC ANHYDRASE - AN OVERVIEW

Carbonic anhydrase (carbonate dehydratase, carbonate hydro-lyase; EC 4.2.1.1) is a zinc containing metalloenzyme that catalyzes the reversible interconversion of $CO_2$ and $HCO_3^-$ and has a maximum turnover number in excess of $10^5$ $s^{-1}$ (Khalifah, 1971). The enzyme was first discovered in human erythrocytes (Meldrum and Roughton, 1933) but has since been found in many organisms including animals, plants, archaebacteria and eubacteria (Hewett-Emmett and Tashian, 1996; Table 1.1). Carbonic anhydrase (CA) plays an important role in many physiological functions that involve decarboxylation or carboxylation reactions, including both photosynthesis and respiration. It also participates in the transport of inorganic carbon ($C_i$) to actively photosynthesizing cells or away from actively respiring cells (Henry, 1996).

The known CAs can be grouped broadly into three independent families (Hewett-Emmett and Tashian, 1996), α-CA, β-CA and γ-CA. These three families have no significant sequence identities, seem to have evolved independently and are an example of convergent evolution of catalytic function (Hewett-Emmett and Tashian, 1996). Although the primary sequences of these CA families are different, the active sites of these three types of CAs contain $Zn^{2+}$ and all of them employ a two-step catalytic mechanism (Lindskog, 1997).

The first step is a nucleophilic attack of a zinc-bound hydroxyl ion on a $CO_2$ molecule residing in a hydrophobic pocket, generating a zinc-bound $HCO_3^-$ (Equation 1). "E" represents the enzyme. The bicarbonate bound to the zinc is replaced by a water molecule, releasing $HCO_3^-$ (Equation 2). The $HCO_3^-$ can gain a $H^+$ to form $H_2CO_3$ or can lose an additional $H^+$ to form $CO_3^{2-}$ (Equation 3). The second step is the regeneration of the active site by the ionization of the

**Table 1.1** A summary of the different families of carbonic anhydrases

| Gene Family | Distribution | Inhibition | Zinc coordination |
|---|---|---|---|
| α | vertebrates, algae, eubacteria, plants and viruses | highly susceptible to sulfonamides | three histidine residues ** |
| β | algae, plants, eubacteria, fungi, invertebrates and archaebacteria | less susceptible to sulfonamides | two cysteine residues and one histidine residue |
| γ | archaebacteria, cyanobacteria, algae* and plants* | highly susceptible to sulfonamides | three histidine residues *** |
| δ | *Thalassiosira weissflogii* (diatom)+ | not known | Sequence does not match a known α-, β- or a γ-CA |

* putative genes exist in EST database ; ** histidine residues are contributed by the same subunit, *** histidine residues are contributed by two different subunits, + CA activity has not been found in the purified protein

water molecule bound to the zinc ion and removal of a proton from the active site. Most carbonic anhydrases have $k_{cat}$ values greater than $10^4$ $s^{-1}$ which requires an intermediate proton shuttle residue (PSR) (Equation 4) to transfer the proton from the metal bound water molecule to the external buffer, "B" (Equation 5). In this step, the zinc ion acts as a Lewis acid to lower the $pK_a$ of the water molecule from ~ 14 to 7.0.

$E\text{-}Zn^{2+}\text{-}OH^- + CO_2 \leftrightarrow E\text{-}Zn^{2+}\text{-}HCO_3^-$ (Equation 1)

$E\text{-}Zn^{2+}\text{-}HCO_3^- + H_2O \leftrightarrow HCO_3^- + E\text{-}Zn^{2}\text{-}H_2O$ (Equation 2)

$HCO_3^- + H^+ \leftrightarrow H_2CO_3 \leftrightarrow 2H^+ + CO_3^-$ (Equation 3)

$PSR + E\text{-}Zn^{2+}\text{-}H_2O \leftrightarrow E\text{-}Zn^{2+}\text{-}OH^- + PSR\text{-}H^+$ (Equation 4)

$PSR\text{-}H^+ + B \leftrightarrow PSR + B\text{-}H^+$ (Equation 5)

Proton transport from the active site is the rate limiting step for enzymes with $k_{cat} > 10^4$ $s^{-1}$. Thus, the $k_{cat}$ is a reflection of the rate of proton transport (Equation 4 and 5), whereas the catalytic efficiency ($k_{cat}/K_m$) is more reflective of the hydration step (Equation 1) and is insensitive to the rate of proton transport. The overall relationship between the three forms of dissolved inorganic carbon ($C_i$) is shown in (Equation 6).

$CO_2 + H_2O \leftrightarrow HCO_3^- + H^+ \leftrightarrow H_2CO_3 \leftrightarrow CO_3^{2-} + 2H^+$ (Equation 6)

The uncatalyzed hydration and dehydration reactions are slow while the dissociation reactions are considered instantaneous. CA accelerates the hydration of dissolved $CO_2$ in solution and increases the rate at which the different forms of $C_i$ interconvert in solution. The equilibrium between the $C_i$ is pH dependent. At physiological ionic strengths, if the pH level is below the first dissociation constant ($pK_1 \approx 6.4$), $CO_2$ predominates; at pH between 6.4 and at about 10.3 ($pK_2$); $HCO_3^-$ predominates; whereas at pH above 10.3, $CO_3^{2-}$ predominates.

In the first part of this chapter, the different types of CAs currently known, their structures and intracellular locations in photosynthetic organisms are discussed. In the second

part, the physiological roles of these CAs in photosynthesis and the $CO_2$ concentrating mechanisms in plants, cyanobacteria and algae have been emphasized along with a brief discussion of photosynthesis, the different CCMs and components of CCMs in plants, cyanobacteria and algae. This dissertation mainly focuses on the different CAs in the green alga *Chlamydomonas reinhardtii*.

**1. α-Carbonic Anhydrases**

The α-type is the most studied CA and is very widely distributed. α-CAs have been found in animals, plants, algae, eubacteria and viruses. α-CA was first identified in human red blood cells. Vertebrate α-CAs can be classified into two groups (Jiang and Gupta, 1999). One group includes the soluble isoforms like CAI, CAII, CAIII, CAIV, CAIX, CAXII and CAXIV. The second group includes membrane associated CAs like (CAIV, CAVI, CAIX, CAXII, CAXIV and CAXVI). Humans have eleven active isozymes including cytoplasmic (CA I, CAII, CAIII and CAVII), mitochondrial (CAVA and CAVB), secreted (CAVI) and membrane associated (CAIV, CAIX, CAXII and CAXIV) forms. Humans also have numerous carbonic anhydrase-related proteins (CA-RPs) that appear not to have CA activity (Sly and Hu, 1995). The range of the specific activities of these isozymes is quite large, with CAII having the highest [2000-3000 WAU/mg; (Wilbur and Anderson unit/mg)] and CAIII (1-5 WAU/mg) the lowest specific activity. All enzymatically active α-CAs have three histidine residues coordinating to the zinc atom (Fig. 1.1). These histidines as well as a number of other residues are well conserved in all active α-CAs. The α-CA structure is dominated by antiparallel β-sheet forming a spherical molecule with two halves. The active site is a funnel-shaped crater with the zinc atom located near the bottom.

In photosynthetic organisms only a few α-CAs have been identified at this time. Most α-CAs are active as monomers of about 30 kDa. One has been identified in each of the

cyanobacteria *Anabaena* and *Synechococcus* (Soltes-Rak *et al*., 1997), and in both cases the CA is localized to the periplasmic space. In contrast, *Synechocystis* PCC6803 does not have an α-CA. The green alga *C. reinhardtii* (Table 1.2) (Karlsson *et al*., 1998) and *Dunaliella salina* (Fisher *et al*., 1996; Yang *et al*., 1999) have three and two α-CAs, respectively. CAs from *D. salina* are located in the periplasm. One of the CA genes of *D. salina* encodes a protein of about 63 kDa that appears to have two active sites, possibly the results of gene duplication and fusion events (Fisher *et al*., 1996).

In *C. reinhardtii*, two of the α-CAs are located in the periplasm and one in the thylakoid lumen. The first α-CA genes cloned from a photosynthetic organism were *Cah1* and *Cah2*, which encode the two periplasmic CAs in *C. reinhardtii* (Fukuzawa *et al*., 1990; Fujiwara *et al*., 1990). These two genes encode very similar proteins although they are differently regulated. *Cah1* is expressed under low $CO_2$ conditions but not under high $CO_2$ conditions. In contrast, *Cah2* is poorly expressed under low $CO_2$ and slightly upregulated under high $CO_2$ conditions. In addition, the expression of *Cah2* under high $CO_2$ appears low compared to the expression of *Cah1* under low $CO_2$ (Fujiwara *et al*., 1990; Rawat and Moroney, 1991). Possibly, *Cah2* resulted from a gene duplication event and has a poorly functioning promoter. Cah1 holoenzyme is a heterotetramer with two 37 kDa subunits and two 4 kDa subunits held together by disulfide bonds (Kamo *et al*., 1990).

The third α-CA gene in *C. reinhardtii, Cah3,* was discovered in 1995 (Karlsson *et al*., 1995). The deduced full length protein contained an amino terminal extension characteristic of bipartite leader sequence. This information along with the immunolocalization studies indicate that this carbonic anhydrase is localized in the lumen of the thylakoid membrane. This α-CA is constitutively expressed and is slightly upregulated under low $CO_2$ conditions (Karlsson *et al*. 1998). Cah3 is a 29 kDa hydrophobic protein with a pI of ~ 7.87. It is not a membrane spanning

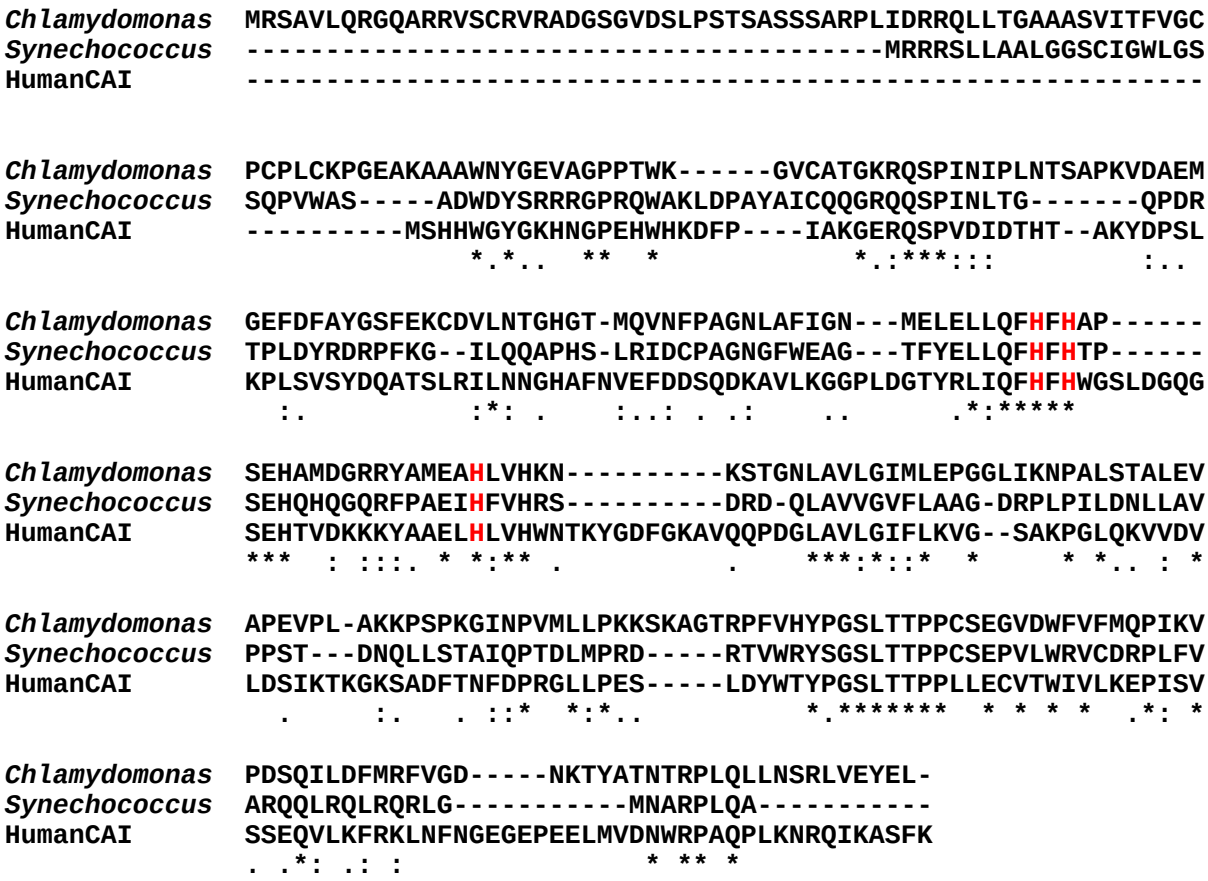



**Figure 1.1 Multiple sequence alignment of α-CAs.** The *C. reinhardtii* CA sequence shown above represents that of thylakoid α-CA, Cah3. Active site residues are shown in bold red. * represents a completely conserved amino acid, **:** represent conserved amino acid substitutions, and **.** represent semi conserved amino acid substitutions.

.

**Table 1.2** A summary of the known CA genes in *Chlamydomonas reinhardtii*

| Gene family | Known gene/s | | References |
|---|---|---|---|
| | Gene name | Location | |
| α | *Cah1*, *Cah2* | Periplasm | Fukuzawa *et al*. 1990; Fujiwara *et al*. 1990 |
| α | *Cah3* | Thylakoid | Karlsson *et al*. 1995 |
| β | *Ca1, Ca2* | Mitochondria | Eriksson *et al*. 1996 |

protein as it can be released into the soluble fraction by washing the thylakoids with 300 mM KCl (Karlsson *et al*.1998). Although the *Cah3* gene has been sequenced, there is no report of the detection and measurement of CA activity of the purified protein. Recently, CA activity of purified recombinant Cah3 was measured and partial biochemical characterization of the purified enzyme was done (Chapter 3 of this dissertation).

It is clear that a number of α-CAs remain to be identified in higher plants. Currently the *Arabidopsis thaliana* database contains six different genes that align with α-CAs (Moroney *et al*., 2001). To our knowledge, only one α-CA cDNA from a higher plant, *ACAH1* from *A. thaliana*, has been completely sequenced (Moroney *et al*., 2001). However a few ESTs in the *Arabidopsis* database do align with some of the other α-CAs identified by the genome project, suggesting that these are expressed genes.

A number of nongreen plant tissues also appear to have CA activity based on the observed high rates of $HCO_3^-$ consumption (Raven and Newman, 1994). Coba de la Pena *et al*. (1997) showed that an α-CA is present in both spontaneously formed and in *Rhizobium meliloti*-induced root nodules of alfalfa. This α-CA is expressed early in the nodule primordium and found later in both developing and mature nodules.

**2. β-Carbonic Anhydrases**

The β-CAs do not appear to be as broadly distributed as the α-CAs at this time. β-CAs were first recognized to be CAs in photosynthetic organisms (Fawcett *et al*., 1990; Burnell *et al*., 1990). Once recognized as a CA, β-CAs have been found in higher plants, algae, eubacteria (Hewett-Emmett and Tashian, 1996), archaebacteria (Smith and Ferry, 1999), the fungi *Saccharomyces cerevisiae* and *S. pombe* (Gőtz *et al*., 1999), *Caenorhabditis elegans* and *Drosophila melanogaster*, but not in vertebrates. In $C_3$ plants, the abundant β-CA is localized to the chloroplast stroma and is a highly active CA. β-CAs have been found both in the cytoplasm

($C_4$ mesophyll cells) and chloroplast ($C_3$ plants) in higher plants and in the cytoplasm of the symbiotic alga *Coccomyxa* (Hiltonen *et al*., 1998).

X-ray absorption spectroscopy studies on spinach β-CA have shown that a histidine and two cysteine residues are the zinc coordinating residues (Bracey *et al*., 1994; Rowlett *et al*., 1994) (Fig. 1.2). Recently the crystal structures of a β-CA from *Porphyridium purpureum* (Mitsuhashi *et al*., 2000) and *Pisum sativum* (Kimber and Pai, 2000) were resolved. The pea CA is an octamer in which dimers form tetramers which form octamers. The red algal CA is double the size of the spinach CA and has two active sites per polypeptide instead of the one found in other β-CAs from algae and higher plants, indicating that a gene duplication event occurred (Mitsuhashi and Miyachi, 1996; Mitsuhashi *et al*., 2000). In the algal CA, an aspartate rather than water occupies the fourth coordination position (Mitsuhashi *et al*., 2000). The red algal CA was crystallized in the absence of substrate (or inhibitor) while the pea CA was crystallized in the presence of acetate. Most probably the difference between the zinc ligands of the two enzymes is due to the conditions under which the crystals were formed. The notion that the $CO_2$ residing in a hydrophobic pocket is required for activity is underscored by the recent comparison between the α-CA and β-CA crystal structures (Kimber and Pai, 2000). It has been found that the three dimensional structure of the active site of the β-CA is a mirror image of that of the active site of the α-CA (Kimber and Pai, 2000).

β-CAs have been found in all photosynthetic organisms studied to date. In the cyanobacterium *Synechocystis* PCC 6803, a β-CA has been described which is encoded by the *icfA* gene. This CA appears to be localized to the carboxysome (Fukuzawa *et al*., 1992; Yu *et al*., 1992). Loss of the carboxysomal β-CA leads to high $CO_2$ requiring phenotype (HCR) (Price and Badger, 1989b). This is the only β-CA in this bacterium as the genome contains no other β-CA gene.

```
Synechococcus  -------------------------------------------------MRKLIEGLR
Coccomyxa      ----------------------------------------MSAKDTADLSPLLEANR
Spinach        TLKEDMAYEEAIAALKKLLSEKGELENEAASKVAQITSELADGGTPSASYPVQRIKEGFI
                                                                :  : *.

Synechococcus  HFRTSYYPSHRDLFEQFAKGQHPRVLFITCSDSRIDPNLITQSGMGELFVIRNAGNLIPP
Coccomyxa      KWADECAAKDSTYFSKVAGSQAPEYLYIGCADSRVSPAQLFNMAPGEVFVQRNVGNLVS-
Spinach        KFKKEKYEKNPALYGELSKGQAPKFMVFACSDSRVCPSHVLDFQPGEAFMVRNIANMVPV
               ::  .   ..   : :.: .* *. : : *:***: *  : :   ** *: ** .*::.

Synechococcus  FG-AANGGEGASIEYAIAALNIEHVVVCGHSHCGAMKGLLKLNQL-QEDMPLVYDWLQHA
Coccomyxa      ---NKDLNCMSCLEYTVDHLKIKHILVCGHYNCGACK--AGLVWH-PKTAGVTNLWISDV
Spinach        FDKDKYAGVGAAIEYAVLHLKVENIVVIGHSACGGIKGLMSFPDAGPTTTDFIEDWVKIC
                      .  :.:**::  *::::::* **  **. *    :         .   *:.

Synechococcus  QATRRLVLDNYSGYETDDLVEILVAENVLTQIENLKTYPIVRSRLFQG-KLQIFGWIYEV
Coccomyxa      REVRDKNAAKLHGLSADDAWDKMVELNVEAQVFNVCASPIVQAAWARGQPLSVHGIVYTP
Spinach        LPAKHKVLAEHGNATFAEQCTHCEKEAVNVSLGNLLTYPFVRDGLVKK-TLALQGGYYDF
                 .:     :  .    :         * ..: *: : *:*:    :   * : *  *


Synechococcus  ESGEVLQISRTSSDDTGIDECPVRLPGSQEKAILGRCVVPLTEEVAVAPPEPEPVIAAVA
Coccomyxa      GTGLVKELIKP---ITGMEDAGALLRADLKQHCF------FSESLA--------------
Spinach        VNG------------SFELWGLEYGLSPSQSV-------------------------
                .*             .::        . .:
```

**Figure 1.2 Multiple sequence alignment of β-CAs.** The full length CAs of *Synechococcus*, *Coccomyxa* and spinach are 272, 227 and 319 amino acids long, respectively. In the figure, 26 amino acids at the carboxyl end of *Synechococcus* CA and 60 amino acids at the amino terminal end of spinach CA are not shown. Active site residues are shown in bold red. * represents a completely conserved amino acid, **:** represent conserved amino acid substitutions, and **.** represent semi conserved amino acid substitutions.

Two nearly identical β-CAs, were identified in 1996 in *C. reinhardtii* (Table 1.2) (Eriksson *et al*., 1996). Both the β-CAs are located in the mitochondria. The molecular weight of the mitochondrial β-CA protein is approximately 21 kDa. These CAs are expressed only under low $CO_2$ conditions (0.035%) and are absent under high $CO_2$ (5%) conditions (Eriksson *et al*., 1998). Recently Giordano *et al*., (2003) have shown that the expressions of these CAs are induced at 0.2% $CO_2$ conditions by increasing the $NH_4^+$ concentration in the growth medium. These workers have proposed that the mitochondrial CAs are involved in providing carbon flux in the form of $HCO_3^-$ to the Tri Carboxylic Acid cycle (TCA) cycle via anaplerosis. The discovery of a third β-CA (*Cah6*) gene in *C. reinhardtii* is reported in Chapter 4.

In *Arabidopsis thaliana*, cDNAs encoding the cytoplasmic and chloroplastic forms of β-CAs have been described (Fett and Coleman, 1994). At this time sequences encoding at least five β-CA genes from *Arabidopsis* are in the database and it is possible that other β-CAs might be found in the future. The $C_4$ monocot plants *Zea mays* (Burnell *et al*., 1999) and *Urochloa panicoides* (Ludwig and Burnell, 1995) also have β-CAs. Translation of the two CA cDNAs in the maize database yields significantly larger CA proteins of about 74 and 60 kDa (Burnell *et al*., 1999). These large polypeptides appear to be a fusion of two monomers since they contain two sets of active site residues. The quaternary structures of the maize CAs have not yet been deciphered. The unusual size of the maize CA protein is unique among higher plants since the cDNAs of $C_4$ plants *U. panicoides* and *Flaveria bidentis* appear "normal" in size and have molecular weights between 24 and 30 kDa. However the fused gene seen in maize is reminiscent of the *Porphyridium* CA (Mitsuhashi *et al*., 2000). Interestingly, all monocot CAs for which sequence information is available lack twelve amino acid residues at the carboxyl terminal end, which are conserved in the dicot CAs and play a role in the oligomerization of the pea CA (Kimber and Pai, 2000).

### 3. γ-Carbonic Anhydrases

A third type of CA, the γ-CA, was discovered in the archaebacterium *Methanosarcina thermophila* (Alber and Ferry, 1994). Genes encoding putative γ-CA proteins have been found in eubacteria and plants (Newman *et al*., 1994). The γ-CA from *M. thermophila* was crystallized and its structure solved (Kisker *et al*., 1996). The structure of the γ-CA is remarkably different from that of α-CA or β-CA. The γ-CA functions as a trimer of identical subunits. Each monomer is a left-handed β-helix (Kisker *et al*., 1996). The trimer contains three zinc atoms, each at the three subunit interfaces. As in α-CAs, three histidines and a water molecule coordinate the zinc atom but the histidines are provided by two subunits (Fig. 1.3). For the *M. thermophila* protein, His 81 and His 122 from one subunit act as ligands and His 117 from a different subunit is the third ligand. In spite of the fact that the active site is at the subunit interface, architecturally the active site of γ-CA resembles that of α-CA (Kisker *et al*., 1996).

A γ-CA homologue, *CcmM* was discovered earlier in *Synechococcus* PCC7942 (Price *et al*., 1993). CcmM lacks detectable CA activity but it is required for growth under low $CO_2$ conditions (Moroney *et al*., 2001). If this gene is mutated, the mutant cannot grow on air levels of $CO_2$, suggesting that it is an essential part of CCM (Price *et al*., 1993). The CcmM protein is over 300 amino acids longer than the *M. thermophila* protein. The N-terminal portion of CcmM has a high homology to the archebacterial γ-CA. The C-terminal portion of CcmM has three to four 87 amino acid repeats that are very similar to the small subunit of Rubisco from the cyanobacteria (Price *et al*., 1998). The exact role of CcmM in the CCM is not clear. The *C. reinhardtii* EST database also has two ESTs that align well with the *M. thermophila* γ-CA. Genomic sequences were obtained for both these gamma CA like protein genes (*Gclp1* and *Gclp2*). One of them, *Gclp1* has been cloned and overexpressed in *Escherichia coli*. A detailed discussion about G*clp1* is in chapter 5 of this dissertation. Several *Arabidopsis* ESTs in the

databases have homology to the γ-CA from *M. thermophila* (Moroney *et al*., 2001). When the databases are searched using sequences around the putative active site of one of these ESTs, three different genomic sequences are obtained. Two of them, one from chromosome I and the other from chromosome V, have very similar sequences around the active site. The third one, a shotgun clone, is also located on chromosome 1 and is much less similar but retains the histidines at the active site (Moroney *et al*., 2001).

**4. Unresolved Questions: Are There More CA Gene Families?**

In 1997, a cDNA encoding a CA from the diatom *Thalassiosira weissflogii* was described by Roberts *et al*. The CA was isolated and a partial amino acid sequence was obtained from the purified protein. Using this protein sequence information, primers were designed to screen a *T. weissflogii* cDNA library by polymerase chain reaction (PCR). The amplified putative CA cDNA sequence was cloned and sequenced. The protein sequence of this CA does not match any known CA sequences. The authors have named the CA family as δ-CA. This raises the possibility that there might be another CA gene family.

Unfortunately the authors could not overexpress the protein in *E. coli* and obtain CA activity. Furthermore, there have been no further reports of CA cDNAs that match the cDNA obtained from *Thalassiosira weissflogii*. Future work is needed to ascertain that there is indeed a δ-CA family, in addition to the other three gene families. Morel's group raised an additional question whether or not the cadmium can substitute for zinc in CA, or if there is a separate cadmium-dependent CA (Lane and Morel, 2000).

Recent work with *Thalassiosira weissflogii* indicates that this diatom might be able to substitute cadmium or cobalt for zinc under zinc limiting environment (Lane and Morel, 2000). The workers demonstrated that the cadmium-or cobalt-CA is different from the CA that they had previously identified in *T. weissflogii* (Roberts *et al*., 1997). This raises the possibility that there

```
Cam   MMFNKQIFTILILSLSLALAGSGCISEGAEDNVAQEITVDEFSNIRENPVTPWNPEPSAP
CcmM  ----------MPSPTTVPVATAGRLAE------------------------------P
                :  . ::.:* :* ::*                                 *

Cam   VIDPTAYIDPQASVIGEVTIGANVMVSPMASIRSDEGMPIFVGDRSNVQDGVVLHALETI
CcmM  YIDPAAQVHAIASIIGDVRIAAGVRVAAGVSIRADEGAPFQVGKESILQEGAVIHGLEYG
       ***:* :.. **:**:* *.*.* *:. .***:*** *: **..* :*:*.*:*.**

Cam   NEEGEPIEDNIVEVDGKEYAVYIGNNVSLAHQSQVHGPAAVGDDTFIGMQAFVFKSKVGN
CcmM  RVLGD---------DQADYSVWIGQRVAITHKALIHGPAYLGDDCFVGFRSTVFNARVGA
      .  *:          *  :*:*:**:.*:::*:: :**** :*** *:*::: **:::**

Cam   NCVLEPRSAAIGVTIPDGRYIPAGMVVTSQAEADKLPEVT-DDYAYS------------
CcmM  GSVIMMHALVQDVEIPPGRYVPSGAIITTQQQADRLPEVRPEDREFARHIIGSPPVIVRS
      ..*:  :: . .* ** ***:*:* ::*:* :**:****  :*  ::

Cam   ----------------------------------HTNEAVVYVNVHLAEGYKETS----
CcmM  TPAATADFHSTPTPSPLRPSSSEATTVSAYNGQGRLSSEVITQVRSLLNQGYRIGTEHAD
                                         :.*.:. *.  * :**:  :
```

**Figure 1.3 Multiple sequence alignment of γ-CAs.** Cam represents the γ-CA from *Methanosarcina thermophila* and CcmM represents the γ-CA from *Synechococcus sp*. Active site residues are shown in bold red. The γ-CA from *Methanosarcina* has an open reading frame of 247 amino acids. The CcmM protein from *Synechococcus* has a 290 amino acids long C-terminal extension. The complete CcmM protein from *Synechococcus* has an open reading frame of 539 amino acids. * represents a completely conserved amino acid, **:** represent conserved amino acid substitutions, and **.** represent semi conserved amino acid substitutions.

can be Cd or Co-requiring CAs in diatoms and also perhaps in algae and higher plants. It has been reported that Co can also substitute for Zn in the γ-CA from *M. thermophila* (Alber *et al*., 1999).

## RIBULOSE 1, 5 BISPHOSPHATE CARBOXYLASE /OXYGENASE (RUBISCO) - AN OVERVIEW

Photosynthesis is the biological process by which most photoautotrophs reduce $CO_2$ to carbohydrates with the release of $O_2$ as a by product of the photochemical oxidation of water. The main enzyme participating in $CO_2$ fixation process is ribulose 1, 5 bisphosphate carboxylase /oxygenase (Rubisco; EC 4.1.1.39 ). In the carboxylation reaction, Rubisco catalyzes the addition of $CO_2$ to ribulose 1, 5 bisphosphate (RuBP), generating two molecules of 3-phosphoglyceraldehyde (3-PGA). Molecules of 3-PGA are converted to glycerladehyde-3-phosphate (GAP) via the Calvin cycle.

Rubisco can also utilize $O_2$ as a substrate. It catalyzes the oxygenation of ribulose 1, 5 bisphosphate, to produce one molecule of 3-PGA and one molecule of 2-phosphoglycolate, a compound that cannot be utilized in the $C_3$ reductive photosynthetic carbon cycle. A second cycle termed the $C_2$ oxidative photosynthetic carbon cycle (photorespiratory carbon oxidation cycle) salvages 75% of the constituent carbon utilizing ATP and NADPH (Hartman and Harpel, 1994). It seems that the oxygenase activity is intrinsic to Rubisco and reflects that the ancestral Rubisco evolved in an anaerobic atmosphere. Thus the relative rates of oxygenation and carboxylation by Rubisco are major factors in determining the efficiency of photosynthesis. Rubisco is catalytically a very slow enzyme with a turnover rate of 3-4 $s^{-1}$ compared to $10^4$ or $10^5$ $s^{-1}$ times for some enzymes like CA and 3-Keto steroid isomerase. The relative rates of catalysis of oxygenation or carboxylation by Rubisco is controlled by the relative concentrations of $O_2$ and $CO_2$ at the site where the enzyme is localized. The chloroplast stroma of mesophyll cells of

higher plants contain about 9 μM $CO_2$ and about 250 μM $O_2$ (Keys, 1986) at 25°C. The $K_m$ values of Rubisco for $CO_2$ and $O_2$ are approximately 8-25 μM and 360-650 μM (Jordan and Ogren, 1983; Keys 1986). As the concentration of $CO_2$ in the stroma of mesophyll cells is approximately equal to the $K_m$ for $CO_2$, Rubisco normally functions at only half of its already-low maximum velocity.

The oxygenation reaction competes with the carboxylation reaction, which slows this carboxylation rate by another 28% under current atmospheric conditions. Given a specificity factor of 100 for Rubisco, the ratio of carboxylase to oxygenase activity is approximately 3 to 1 under present atmospheric conditions. Thus to sustain reasonable rates of photosynthetic $CO_2$ fixation, plants must make an enormous amount of Rubisco. The enzyme may constitute up to 25% of the total nitrogen in a plant and up to 50% of the protein in the chloroplast stroma. Considering the quantity of photosynthetic tissue, Rubisco is probably the most abundant enzyme in the biosphere (Goodwin and Mercer, 1983).

In higher plants, cyanobacteria and most eukaryotic algae, Rubisco is composed of eight large subunits and eight small subunits. The holoenzyme has eight active sites. Each active site of Rubisco is catalytically active only after the formation of a ternary complex with $CO_2$ and $Mg^{2+}$. This reaction, called activation, begins with the covalent binding of a $CO_2$ molecule to the ε-amino group of a lysine residue in the large subunit. The resulting carbamate is then stabilized by the binding of $Mg^{2+}$ (Schneider *et al*., 1992). The $CO_2$ molecule that forms the carbamate is termed the activator $CO_2$ and is distinct form the substrate $CO_2$ molecule. In addition to stabilizing the carbamate, the metal ion plays an important role in some aspect of catalysis and may interact with the gaseous substrate. This is based on the finding that other divalent metal ions like $Ni^{2+}$, $Co^{2+}$, $Fe^{2+}$, $Mn^{2+}$ and $Cu^{2+}$ can substitute for $Mg^{2+}$ in stabilizing carbamate and the catalytic rates for these forms of the enzyme are low and differ in their ability to discriminate

between $CO_2$ and $O_2$ (Andrews and Lorimer, 1987). $Mg^{2+}$ binding may promote the proper binding of the phosphate groups of RuBP.

The reaction mechanism is strictly ordered. RuBP binds first and then reacts with either $CO_2$ or $O_2$. After binding of RuBP to an activated site, the first step in catalysis is the deprotonation of C3 to form a 2, 3-enediol, which creates a nucleophilic center at C2. At this point $CO_2$ and $O_2$ compete for the nucleophilic center at C2, with carboxylation producing two molecules of 3PGA and oxygenation producing one molecule of 3 PGA and one molecule of 2-phosphoglycolate. Because of the catalytic properties of Rubisco, $CO_2$ and $O_2$ are competitive substrates and an increase in the $CO_2$ concentration favors the carboxylation rate, leading to the effective fixation of $CO_2$ by Calvin cycle.

## FUNCTIONAL ROLES OF DIFFERENT CARBONIC ANHYDRASES IN THE PHOTOSYNTHETIC CARBON DIOXIDE CONCENTRATING MECHANISM (CCM)

Higher plants, aquatic angiosperms, cyanobacteria and algae have all developed their own unique versions of photosynthetic CCMs to aid Ribulose 1, 5 bisphosphate carboxylase /oxygenase (Rubisco) in efficient $CO_2$ capture as discussed in the earlier sections. An important aspect of all CCMs is the critical roles that the various intracellular and extracellular CAs play in the CCM. In the following section emphasis will be placed on the functional roles of CAs in different photosynthetic CCMs along with an overview of photosynthesis and different CCMs that exist in plants, cyanobacteria and algae.

### 1. Effect of $CO_2$ and Presence of a Carbonic Anhydrase on Photosynthesis

Ambient air contains about 0.036% $CO_2$ and 21% $O_2$. In terrestrial plants, $CO_2$ from the atmosphere enters the leaves through the stomata and diffuses through the substomatal chamber, the intercellular spaces and enters into the mesophyll cells to reach the carboxylation site of Rubisco, which is located in the stroma of the chloroplast. Since $CO_2$ has to diffuse from the

external environment to reach the carboxylation site, the efficiency of carboxylation depends on the diffusion rate of $CO_2$. The rate of $CO_2$ diffusion is $10^4$ times slower in water than in air. As stated earlier, the chloroplast stroma of mesophyll cells of higher plants contain about 9 μM $CO_2$ and about 250 μM $O_2$ (Keys, 1986) at 25°C. In aquatic environments, the concentrations of $CO_2$ and $O_2$ can vary from 0 to 250 μM and 0 to 500 μM, respectively (Bowes and Salvucci, 1989). These estimates underscore the variation in the concentrations of $CO_2$ faced by aquatic photoautotrophic organisms.

Studies in different species of higher plants have revealed that the $C_3$ plants had significant increases in dry mass when the concentration of $CO_2$ was increased compared to the $C_4$ plants grown under the same conditions (Akita and Tananka, 1973). This indicates that many photoautotrophic organisms operate under conditions in which their photosynthetic rates are limited by the rate of carboxylation reaction. This suggests that the efficiency of carbon fixation during photosynthesis may be improved by a Rubisco that is catalytically more active (Andrews and Lorimer, 1987), by increasing the substrate specificity of Rubisco for $CO_2$ or by elevating the concentration of $CO_2$ at the site of Rubisco activity.

Most studies have shown that since the control of photosynthesis is shared by several components of the system under ambient conditions, simply increasing the catalytic activity of Rubisco would probably not increase the photosynthetic rate. The specificity for $CO_2$ does vary among different species. The specificity for $CO_2$ is higher in $L_8S_8$ forms than in $L_2$ forms and is highest in Rubisco from higher plants. Intense selection pressure and extensive mutagenesis experiments have failed to produce a more discriminant form of the enzyme. This suggests that modifications to eliminate or reduce oxygenations involve multiple changes in the enzyme which were not favored by evolutionary changes and would not occur in mutagenesis experiments (Andrews and Lorimer, 1987). The fact that inorganic carbon concentrating mechanisms have

evolved at different instances, implies that these mechanisms probably provide a most cost-effective means of improving photosynthetic rates.

β-CA is present in the chloroplast of mesophyll cells in $C_3$ plants. Potential roles of the β-CA present in the chloroplast of $C_3$ plants range from modulation of the pH of stroma to facilitating diffusion of $CO_2$ across the chloroplast envelope. It is also proposed that the β-CA replenishes the $CO_2$ supply in the stroma from $HCO_3^-$ which is more abundant in the alkaline stroma.

Transgenic tobacco plants were made containing antisense CA constructs by two different groups, to study the function of the β-CA. One group reported no deleterious effects of the reduction of CA activity in the mutant (Price *et al*., 1994). The other group reported that the antisense plants compensated for the decrease in CA with an increase in stomatal conductance leading to an increase in water loss (Majeau *et al*., 1994). The reasons for the drought sensitivity are not clear although it is possible that a reduction in photosynthesis due to a decrease in the delivery rate of $CO_2$ to Rubisco caused the leaf stomata to remain open. It is also possible that lowering the chloroplast β-CA somehow resulted in a disruption of the signal for the plant to close the stomata under certain conditions.

Badger and Pfanz, in 1995, demonstrated the crucial dependence of $C_4$ photosynthesis on the β-CA activity in the mesophyll cells by using CA inhibitors. In *Arabidopsis* there is a report that shows that the β-CA is necessary for photosynthesis. Antisense transgenic *Arabidopsis* plants showing 90% reduction of the β-CA activity were generated by Hee Jin Kim (1997). These antisense CA plants resembled wild type plants when grown either under high $CO_2$ (2%) conditions or on Murashige-Skoog (MS) containing 2% sucrose (Hee Jin Kim, 1997). When grown in ambient $CO_2$ conditions (0.03%) or on sucrose free media, most antisense CA plants died. It is possible that the chloroplast β-CA played a more critical role in photosynthesis

during the early Cretaceous era when $CO_2$ concentrations were lower than they are today. The increase of $CO_2$ in the atmosphere may have rendered the chloroplast β-CA expendable except when $CO_2$ is limiting.

**2. $CO_2$ Concentrating Mechanism (CCM)**

A number of photosynthetic organisms have developed mechanisms to increase the level of $CO_2$ at the location of Rubisco, minimize the energy consuming deleterious oxygenation reaction and reduce nitrogen allocation cost in the form of Rubisco. These include $C_4$ photosynthesis (Hatch, 1987) and Crassulacean Acid Metabolism (CAM), seen in a number of higher plant families and a $CO_2$ concentrating mechanism seen in some microalgae (Osterlind, 1950; Berry *et al*., 1976) and cyanobacteria (Turpin *et al*., 1984; Price *et al*., 1992).

These photosynthetic organisms package Rubisco in a very specific location, have novel means to concentrate $CO_2$ and/or $HCO_3^-$, have a means of rapidly converting the accumulated $HCO_3^-$ to $CO_2$ if $HCO_3^-$ is concentrated and have mechanisms to deliver the $CO_2$ to the location of Rubisco. Environments habituated by many photosynthetic bacteria as well as micro and macro algae, show great fluctuations in concentrations of gases. Many of these organisms employ a mechanism of concentrating $CO_2$ that is induced only under low $CO_2$ conditions (0.036% $CO_2$ in air) and does not operate when high $CO_2$ conditions are prevalent (Osterlind, 1950; Berry *et al*., 1976). A $C_4$ type mechanism has not been discovered in these organisms and the Rubisco from these organisms have a high $K_m$ for $CO_2$ (Jordan and Ogren, 1981) implying the presence of a novel $CO_2$ concentrating mechanism. The CCM provides these organisms with another additional ecological advantage other than that shared by terrestrial $C_4$ and CAM plants. CCM allows both $CO_2$ and $HCO_3^-$ species to be efficiently exploited. The utilization of $HCO_3^-$ is an advantage as in neutral and alkaline environments $HCO_3^-$ is the most abundant form of inorganic carbon ($C_i$).

All the $CO_2$ concentrating mechanisms consist of the following components (Badger, 1987):

a) A means to transport inorganic carbon species against a concentration gradient across either the plasma membrane or the chloroplast envelope or both.

b) A source of energy supply to drive the uphill inorganic carbon transport.

c) A compartment where Rubisco is localized separately from the reactions that accumulate the intermediate Ci pool. For $C_4$ plants this is the bundle sheath, for cyanobacteria it is the carboxysome and for algae it appears to be the pyrenoid.

d) A means for releasing $CO_2$ from the captured pool, co-localized with Rubisco in the special compartment. This might be decarboxylating enzymes like $NAD^+$/$NADP^+$ malic enzyme as in $C_4$ plants and CAs in cyanobacteria and algae.

e) A mechanism to prevent $CO_2$ efflux from the site of Rubisco to ensure efficient $CO_2$ fixation.

**3. Cyanobacterial $CO_2$ Concentrating Mechanism**

Cyanobacteria (blue green algae) have evolved a remarkable environmental adaptation for survival at limiting $CO_2$ conditions. This adaptation is known as a $CO_2$ concentrating mechanism (CCM). A distinguishing feature of the cyanobacterial CCM relative to CCMs in other aquatic organisms is the existence of a constitutive form of CCM in cells grown even at hyper-normal levels of $CO_2$ (2-5% of $CO_2$). These high $C_i$ cells have the ability to use $CO_2$ and $HCO_3^-$ as substrates and to accumulate significant levels of $C_i$ with a relatively high photosynthetic affinity for $C_i$ transport.

Cyanobacterial cells exposed to $C_i$ limitation (typically 20-50 ppm $CO_2$) and air (350 ppm $CO_2$) have the ability to express an enhanced level of CCM activity. This change is accompanied by an increase in Rubisco activity (Price *et al*., 1992), about a two fold increase in carboxysome content (McKay *et al*., 1993; Turpin *et al*., 1984), transcriptional upregulation of

transport activities and increases in the affinities for $CO_2$ and $HCO_3^-$ uptake activities (Badger and Price, 1992; Kaplan and Reinhold 1999; Kaplan *et al*., 1994; Price *et al*., 1998). Typically a 20 fold decrease in the $K_m$ ($C_i$) is seen when cells are grown at 20-30 ppm $CO_2$ levels and more than a half of this rise in affinity is due to an increase in the affinity for $HCO_3^-$ uptake (Sültemeyer *et al*., 1995; Yu *et al*., 1994).

Most of the studies of physiological and molecular aspects of the cyanobacterial CCM have been done on the freshwater cyanobacterium *Synechococcus* PCC7942. This is largely because these strains are easily grown in the laboratory under the desirable defined conditions of rapid aeration and high light and their excellent suitability for genetic modifications using recombinant technologies. With the availability of the *Synechosystis* PCC6803 genomic database, considerable attention has been turned to this species as a model for CCM studies. Basic features of the cyanobacterial CCM include the following (Fig. 1.4):

a) Existence of at least four distinct modes of active uptake for $CO_2$ and $HCO_3^-$, with two transport activities being constitutive and another two being inducible by growth under $C_i$ limitation.

b) Irrespective of the type of uptake, there is a resultant accumulation of $HCO_3^-$ within the cells that can be as high as thousand fold with respect to total exogenous levels.

c) Accumulated $HCO_3^-$ is used to increase the $CO_2$ concentration at the site of Rubisco, which is encapsulated in a unique micro-compartment called the carboxysome.

d) The carboxysome contains a specific carbonic anhydrase for conversion of $HCO_3^-$ to $CO_2$ at a rate high enough to match the maximal rate of $CO_2$ fixation.

e) Existence of an effective mechanism for minimizing the leakage of $CO_2$ from the site of elevation in the carboxysome. This mechanism is supplemented by an efficient $CO_2$ uptake mechanism that can recycle leaked $CO_2$ back into the cell as $HCO_3^-$.

**a. Cyanobacterial Transport Systems for $C_i$**

The model cyanobacterium *Synechococcus* PCC7942 possesses at least four distinct modes of $C_i$ uptake when grown under $C_i$ limitation, each possessing a high degree of functional redundancy. Two of these modes of Ci uptake are induced under low $CO_2$ conditions and the remaining two are constitutively expressed. These four $C_i$ uptake modes are:

a) The inducible high affinity, $Na^+$ independent $HCO_3^-$ transporter, BCT1 was the first cyanobacterial $C_i$ uniporter to be identified and characterized. In *Synechococcus sp.*, BCT1 is encoded by the *cmpABCD* operon and is expressed under $C_i$ limitation (Omata *et al*., 1999). The deletion mutant, *cmpAB*$^-$ (M42) has a low affinity for $HCO_3^-$ (Okamura *et al*., 1997). BCT1 is a member of the diverse subfamily of bacterial ABC (ATP binding cassette) (Higgins, 1992) transporters. The *cmpA* codes for the precursor of the 42 kDa, $HCO_3^-$ binding protein (cmpA) which is closely related to NrtA (Omata, 1991), a nitrate/nitrite binding lipoprotein for the nitrate/nitrite transporter (*nrtABCD*) in *Synechococcus* PCC7942.

b) Cells of *Synechococcus* PCC7942 also possess a $Na^+$-dependent $HCO_3^-$ uptake activity. It has been suggested that *Synechococcus* PCC7942 cells may possess a $Na^+/HCO_3^-$ symporter that is energized by inwardly directed $Na^+$ gradient (Espie and Kandasamy, 1994). An indirect role of $Na^+$ through a need for pH regulation via $Na^+/ H^+$ antiport mechanism is also possible. A gene (*ictB*) has been suggested as a candidate for constitutive $Na^+$-dependent $HCO_3^-$ transporter as a result of analysis of a HCR mutant in *Synechococcus* that harbors a complex tag mutation. The gene codes for a hypothetical membrane protein with nine to ten membrane spanning domains. Another HCR mutant in *Synechococcus* has a lesion in the gene *dc14* (Ronen-Tarazi *et al*., 1998), which encodes a putative $Na^+$-dependent $HCO_3^-$ transporter. This transporter may be involved in a fast induction response to low $CO_2$ in cyanobacteria (Sültemeyer *et al*., 1997). Recently a gene from *Synechocystis* PCC6803 has been isolated that appears to code for a $Na^+$-

dependent bicarbonate transporter (Shibata *et al*., 2002). The gene is known as *sbtA* (SLR 1512) and a mutant with lesions in this gene is unable to grow under low $HCO_3^-$ at pH 7 or 9 (Shibata *et al*., 2002). The sbtA protein has 8 to 10 membrane spanning domains and a hydrophobic domain near the center of the protein that may represent a membrane extrinsic region.

c) and d) There are two $CO_2$ uptake systems based on specialized forms of thylakoid-based Type 1 NADPH dehydrogenase complexes (NDH-1). One of the $CO_2$ uptake systems is constitutive and uses a specialized form of the thylakoid based, Type 1 NADPH complex referred to as NDH-$1_4$, with specific subunits coded by *NDHF4*, *ChpX* and *NDHD4*. The other $CO_2$ uptake system referred to as NDH-$1_3$, is induced under low $CO_2$ conditions and has subunits coded by *NDHF3- NDHD3-ChpY* operon. The two unique proteins ChpX and ChpY (also called CupA and CupB) (Maeda *et al*., 2001; Shibata *et al*., 2001) have $CO_2$ hydration activity in the light. They have no sequence homology with the known families of CA proteins. However, ChpX has two conserved histidine and one conserved cysteine residues in the highly conserved region in the protein. These conserved residues could be potential coordination sites for a Zn atom in the active site. ChpX is bound to the NDH-1 complex and is in close contact with NDHF4 and NDHD4 so as to create a "proton wire" for abstraction of protons to the thylakoid lumen via the photosynthetic electron transport chain. ChpY can function in a similar way (Price *et al*., 2003). Thus $C_i$ uptake is tightly light dependent.

**b. Rubisco Localization and the Role of the Carboxysome in CCM in Cyanobacteria**

In cyanobacteria (blue green algae), a number of alterations in the structure, localization and conformation of Rubisco have been shown to cause HCR mutants. Immunogold labeling shows Rubisco is localized to the micro compartment called the carboxysome (McKay *et al*., 1993). Carboxysomes are proteinaceous, semicrystalline bodies surrounded by a proteinaceous shell. $C_i$ is delivered into the cell in the form of $HCO_3^-$ and remains largely as $HCO_3^-$ except in

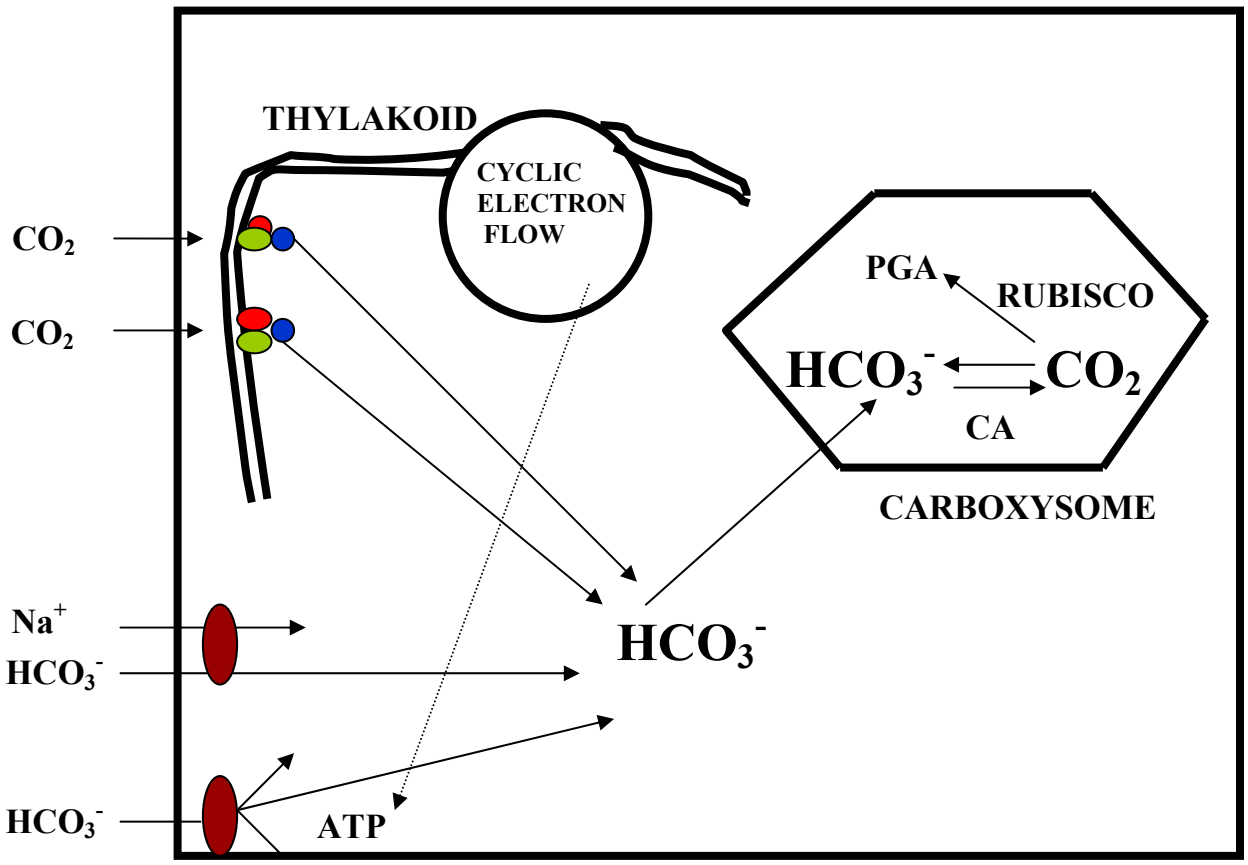



**Figure 1.4 A proposed model for CCM in cyanobacteria.** The font sizes of $CO_2$ and $HCO_3^-$ indicate the relative concentration of these Ci species.

the carboxysome where a β-CA converts the $HCO_3^-$ to $CO_2$ (Badger and Price, 1992). This increases the concentration of $CO_2$ at the site of Rubisco, ensuring efficient $CO_2$ fixation.

The carboxysomal CA has a 60-70 amino acid extension at the carboxyl end of the protein (Badger and Price, 1994) when compared with other prokaryotic β-CAs giving a monomer size of 31 kDa compared to 24 kDa for the chloroplast enzyme of higher plants. The proximal part of the carboxyl extension may be required for the oligomerisation of the CA and this oligomerization is essential for catalysis (So *et al*., 2002).

Recent genome analysis of seven cyanobacterial genomes has lead to the identification of two types of carboxysomes in cyanobacteria (Badger, 2003). One type has the typical cyanobacterial carboxysomes and the other type has carboxysomes similar to autotrophic β-proteobacteria like *Thiobacillus* species. In one marine cyanobacterium *Synechococcus* WH8102, the CA does not have the carboxyl terminal extension found in other carboxysomal CAs (Badger *et al*., 2002). Moreover, *Prochlorococcus marinus* MED4 and MIT9313 have no identifiable CAs in their genome (Badger *et al*., 2002). This clearly raises questions about the presence of carboxysomal CA in this type of carboxysome and also about the roles of CAs in general in these species. It also asks if yet to be identified CAs, are present in these organisms.

To prevent the dissipation of the $HCO_3^-$ pool, CA activity should be absent in the cytosol of cyanobacteria. The expression of human CA in the cytoplasm of *Synechococcus* PCC7942 cells results in a massive leakage of $CO_2$ from the cells producing a HCR phenotype and implicating the carboxysomes as the site of $CO_2$ elevation (Price and Badger, 1989a). Also the cyanobacterial genome sequences do not appear to code for a CA that is expressed in the cytosol. (Badger *et al*., 2002). A cluster of five genes, *ccmKLMNO*, have been shown to be involved in structural assembly or functioning of carboxysome. Mutations in some of these genes (*ccmA* and *ccmJ*) produce mutant cells that have HCR phenotypes and are unable to utilize the intracellular

$C_i$ pool although they are capable of $C_i$ transport (Badger and Price, 1994; Ronen-Tarazi, *et al*., 1995).

The *CcmM* gene codes for the γ-CA analog, CcmM. Cells deleted in *CcmM* show HCR phenotype and have empty carboxysomes. From these results it is clear that CcmM is required for correct carboxysome assembly and for optimal growth on low levels of $CO_2$. It is not clear, if CcmM has CA activity or its enzymatic activity is needed for correct assembly of carboxysomes. $CO_2$ leakage out of the carboxysome has been hypothesized to be prevented by the spatial arrangement of CA and Rubisco in the carboxysomes (Reinhold *et al*., 1991), which ensures that the $CO_2$ produced is fixed before it can leak to the cytoplasm. $CO_2$ leakage is further minimized by the carboxysomal proteinaceous shell which is selectively permeable to $HCO_3^-$ and not to $CO_2$.

Characterization of a *Synechococcus* mutant containing a disruption in an ORF (ORF 272) led to the isolation of the carboxysomal β-CA gene *icfA* and its product (Fukuzawa *et al*., 1992; Yu *et al*., 1992). The loss of this CA leads to a cell that cannot grow very well on limiting levels of $CO_2$. This mutant accumulates $HCO_3^-$ to higher levels than wild type cells, presumably because these cells can no longer convert $HCO_3^-$ to $CO_2$. This CA is co-localized in the carboxysome with Rubisco (Price *et al*., 1992).

Replacement of Rubisco of *Synechocystis* PCC6803 with that of *Rhodospirillum rubrum*, led to the loss of carboxysomes and the mutant showed a HCR phenotype, although it was capable of accumulating $C_i$. Studies indicated that this *Rhodospirillum* Rubisco was not packaged into the carboxysome. *Rhodospirillum* Rubisco lacks small subunits and has poor specificity for $CO_2$ versus $O_2$. In cyanobacteria the small subunits might play a role in the structural organization of the carboxysome.

Characterization of one cyanobacterial HCR mutant, EK6, showed that it contains a thirty amino acid extension in the Rubisco small subunit (Orús *et al*., 1995). This resulted in a Rubisco

that was incapable of packing into the carboxysome. *In vitro* analysis showed that this Rubisco has the same $K_m$ ($CO_2$) as the wild type enzyme (Schwarz *et al*., 1995). Another HCR mutant, Mu28 (Friedberg *et al*., 1993), does not have any apparent carboxysomes and immunolabelling studies showed that the Rubisco is distributed throughout the mutant cell while in wild type cell it was localized in the carboxysome.

**c. Other Genes and Proteins Involved in the Cyanobacterial CCM**

A number of other genes affect the $C_i$ accumulation mechanism in cyanobacteria, apart from genes contributing to Rubisco, carboxysome and NADH dehydrogenase complex (Price *et al*.,1998). These genes include:

1) *RbcR* There are two copies of the putative transcription regulator gene *RbcR* (*Rbc*R1 and *Rbc*R2) in *Synechocystis* PCC6803 genome. Loss of induction of the *cmp* operon occurs at low $CO_2$ conditions when *Rbc*R1 is inactivated. This indicates that the *Rbc*R1 regulates transcription under $C_i$ limitation.

2*) icfG* In *Synechocystis*, the gene *icfG* has been implicated in the down regulation of $C_i$ uptake ability. *icf*G is expressed in the presence of glucose and is required for the switch from photoautotrophic to photoheterotrophic growth (Beuf *et al*., 1994).

3) *SLR0143* This gene codes for a large protein with similarity to a beta transducin like protein Mutation in this gene results in HCR phenotype with reduced affinity for $C_i$ uptake (Bédu *et al*., 1995).

4) *cotA* It codes for a membrane protein that has a haem binding site similar to chloroplast envelope protein *Cem*A (Katoh *et al*. 1996). *Cot*A mutants have reduced $C_i$ uptake and reduced $H^+$ extrusion during initial uptake phase. This mutant does not grow at low $Na^+$ conditions at neutral pH but can grow at high pH even at low $Na^+$. It is speculated that this protein helps in proton extrusion from the cell during $HCO_3^-$ uptake.

5) *ecaA* This gene codes for a periplasmic α-CA in *Synechococcus sp.* and *Anabaena* PCC7120 (Soltes-Rak *et al*., 1997). This gene is upregulated under high $CO_2$ conditions and down regulated under low $CO_2$ conditions. Insertional inactivation of this gene did not produce a clear phenotype so whether or not the activity of this α-CA facilitates the supply of $CO_2$ to the plasma membrane remains unclear. It is likely that it facilitates the diffusion of $CO_2$ across the plasma membrane. *Synechocystis* PCC6803 lacks this α-CA.

**4. $CO_2$ Concentrating Mechanism in *Chlamydomonas***

The ability of unicellular green algae to grow well both under low and high $CO_2$ conditions was first discovered by Osterlind (1950). It has been proposed that unicellular green algae grown under limiting levels of $CO_2$ acquire the ability to utilize $CO_2$ more efficiently (Berry *et al*., 1976). A specific inducible CCM has been suggested to explain this phenomenon (Badger *et al*., 1980; Aizawa and Miyachi, 1986; Badger and Price, 1992). This low $CO_2$ inducible CCM does not operate in high $CO_2$ grown cells. The main function of this CCM is to elevate the $CO_2$ concentration at the site of Rubisco. Ci accumulation has been observed in many algal species including red algae (Burns and Beardall, 1987), diatoms (Coleman and Rotatore, 1995), marine green algae (Burns and Beardall, 1987; Beer *et al*., 1990) fresh water algae symbiotic algae (zooxanthellae) and dinoflagellates (Berman-Frank *et al*., 1998; Leggat *et al*., 1999). When grown under high $CO_2$ conditions (3%-5% $CO_2$ in air) these organisms show low affinity for $CO_2$ and high rates of photorespiration. Under low $CO_2$ conditions (0.035%), these organisms show high affinity for $CO_2$, low rates of photorespiration and an ability to concentrate $C_i$ at a level higher than that in the surrounding environment. This CCM mechanism is different from the $C_4$ mechanism of $CO_2$ concentration. The algal CCM is different from the cyanobacterial CCM, probably due to the reason that the prokaryotic cells have one set of limiting membranes while the eukaryotic green algae have two sets of membranes for $C_i$ to cross.

**a. The Proposed Model**

Several CCM models have been proposed (Moroney *et al*., 1986; Moroney and Mason, 1991; Badger and Price 1994). Moroney and Mason (1991) proposed the model shown in Fig. 1.5. According to this model, a CA-promoted diffusion of $CO_2$ occurs across the plasma membrane as seen in *C. reinhardtii* and *Chlorella saccharophila* strains. Under low $CO_2$ conditions, these algae accumulate high concentrations of periplasmic CA (Kimpel *et al*., 1983). $CO_2$ diffuses across the plasma membrane into the cytoplasm. In the cytoplasm, this $CO_2$ is converted into $HCO_3^-$ by either an uncatalyzed reaction or in the presence of a cytoplasmic CA. This $HCO_3^-$ in the cytoplasm is then actively transported into the chloroplast. In the chloroplast, $HCO_3^-$ diffuses to the pyrenoid. The pyrenoid is a large protein complex surrounded by a sheath of carbohydrates like starch, amylose or paramylon (Gibbs, 1962; Okada, 1992). The majority of the Rubsico is localized in the pyrenoid in algae. Another CA, which may be located in the chloroplast envelope or thylakoid or in the pyrenoid, catalyzes the conversion of $HCO_3^-$ to $CO_2$. This $CO_2$ can diffuse to the site of Rubisco. This elevation of $CO_2$ at the site of Rubisco reduces the oxygenation reaction of Rubisco.

**b. Physiological Studies**

Evidence for $CO_2$ as the inorganic carbon species crossing the plasma membrane in *C. reinhardtii* comes from experiments using $C_i$ as a substrate for fixation studies (Tsuzuki, 1983), the effect of external pH on $CO_2$ fixation (Moroney and Tolbert, 1985) and from the use of impermeant inhibitors of CAs like acetazolamide (AZ) (Moroney *et al*., 1985). $CO_2$ fixation and $C_i$ accumulation were inhibited by AZ (Moroney *et al*., 1985). This shows $HCO_3^-$ is not the Ci species transported across the plasma membrane. Organisms growing in acidic conditions like *C. saccharophila*, can take up only $CO_2$ while those growing in alkaline conditions can take up both $CO_2$ and $HCO_3^-$ (Beardall, 1981). Active uptake of $HCO_3^-$ has been suggested in the green alga

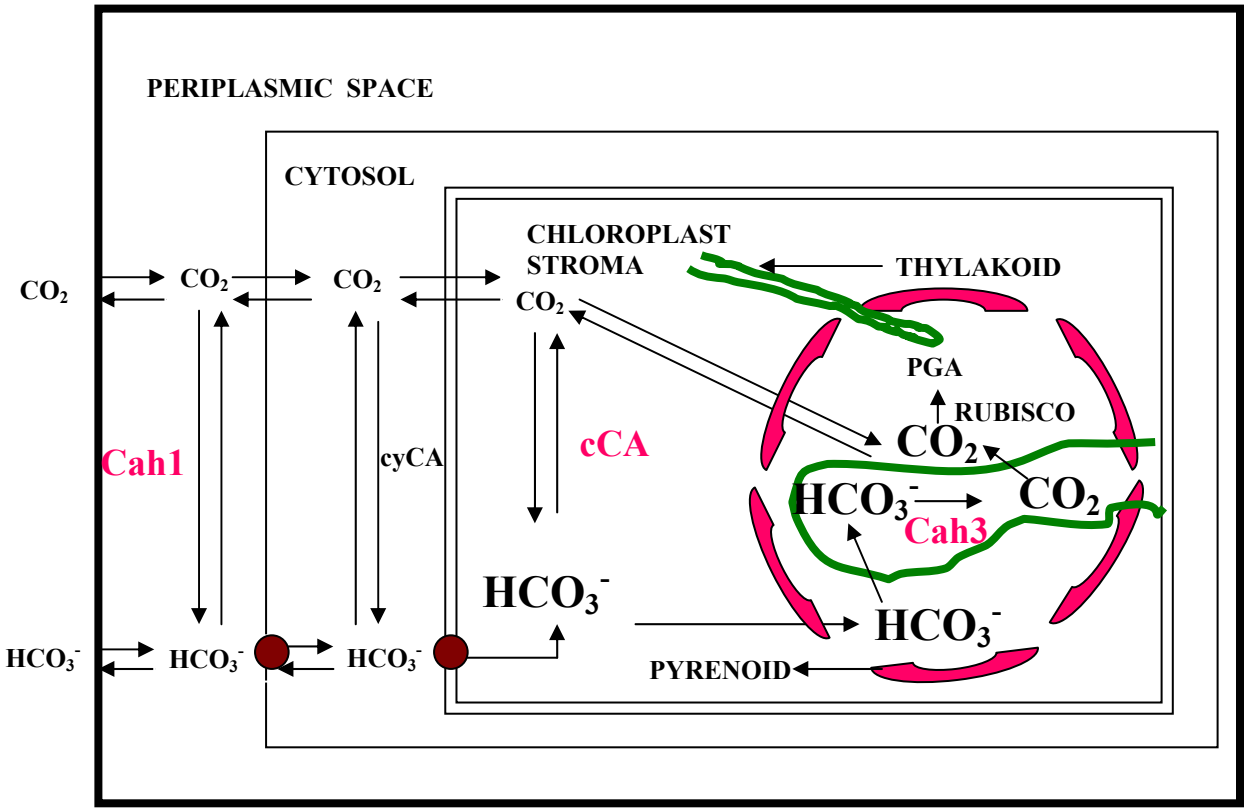



**Figure 1.5 A model showing the potential roles of carbonic anhydrases in the operation of CCM in *C. reinhardtii*.** The font sizes of $CO_2$ and $HCO_3^-$ indicate the relative concentration of these Ci species. cCA and cy CA represent chloroplastic and cytoplasmic carbonic anhydrases. Cah1 and Cah3 represent the periplasmic and thylakoid CA, respectively. Putative $HCO_3^-$ transporters are denoted by brown small circles.

*Scenedesmus* (Thielmann *et al*., 1990) because $C_i$ transport in this alga is not affected by AZ. This indicates that the organism does not need a CA in the periplasm.

An active transport has also been proposed in *C. reinhardtii* (Sültemeyer *et al*., 1989). Transport activities for $CO_2$ have been found on the chloroplast envelope (Amoroso *et al*., 1998) and there is also evidence for $HCO_3^-$ uptake across the cell membrane (Palmqvist *et al*., 1994; Amoroso *et al*., 1998). However, to date, $CO_2$ and/or $HCO_3^-$ transporters at the plasma membrane or chloroplast envelope have not been identified. The NdhF, NdhD and Chp subunits associated with the NDH-1 complex in blue green algae (cyanobacteria) have no homology in the EST database of *C. reinhardtii*. The argument that active transport at the plasma membrane is not the main transport step for $C_i$ is strengthened by the evidence that isolated chloroplasts can concentrate $CO_2$ at low $CO_2$ conditions (Moroney *et al*., 1987; Goyal and Tolbert, 1988). When grown under high $CO_2$ conditions, green algal cells behave like $C_3$ plants, producing large amounts of photorespiratory products and showing a $K_m$ ($CO_2$) of 20 μM (Moroney and Tolbert, 1985). On acclimation to low $CO_2$ conditions, the $K_m$ ($CO_2$) decreases to 2 μM along with a significant reduction of photorespiratory byproducts. Low $CO_2$ grown cells accumulate acid-labile $^{14}C$, in the presence of light and $NaH^{14}CO_3$. This implies that there is an increase in the concentration of intracellular $C_i$. $C_i$ accumulation can be blocked by inhibitors of electron transport or photophosphorylation (Badger and Price 1994; Berry *et al*., 1976). This indicates that light energy establishes and maintains this inorganic carbon concentrating mechanism.

**c. Low $CO_2$ Inducible Proteins and Genes in *Chlamydomonas reinhardtii***

When CCM is induced under low $CO_2$ conditions, the algal cells preferentially synthesize some new proteins from newly transcribed mRNAs. Based on the inducibility of CCM, two approaches have been employed to identify the components involved in this mechanism. One method involves the identification, purification, sequencing and characterization of the

newly synthesized polypeptides. The second involves the construction of a cDNA library using mRNA from low and high $CO_2$ adapted cells. These cDNA libraries could then be differentially screened for cDNAs induced or unregulated under low $CO_2$ conditions. A description of some of these proteins and genes are given below.

Based on physiological studies, it can be assumed that the proteins that are involved in CCM are induced or derepressed only on adaptation to low $CO_2$ conditions (Badger *et al*., 1980). *In vivo* labeling studies using $^{35}SO_4^{2-}$ demonstrated that five polypeptides of molecular weights 46, 44, 37, 36 and 20 kDa are preferentially synthesized in *C. reinhardtii* under low $CO_2$ conditions (Manuel and Moroney, 1988). The 37 kDa protein has been identified as a periplasmic CA (Cah1) (Coleman and Grossman, 1984; Fukuzawa *et al*., 1990). There are two periplasmic α-CA genes, namely *Cah1* and *Cah2*. Cah1 is required for the functional operation of CCM in *C. reinhardtii* (Moroney *et al*., 1985). Cah1 facilitates the diffusion of both $CO_2$ and $HCO_3^-$ to the cell surface by promoting rapid equilibrium. $CO_2$ is then available for both passive diffusion and active uptake across the cell membrane and $HCO_3^-$ is available for $HCO_3^-$ uptake systems. The concept that $HCO_3^-$ is the main Ci species that is transported across the plasma membrane is supported by the observation that CA inhibitor AZ inhibits $CO_2$ fixation at high pHs where $HCO_3^-$ is predominant (Moroney *et al*., 1985). Recent studies have cast some doubt on these concepts when a *Chlamydomonas* mutant defective in *Cah1* gene appeared to perform as well as wild type cells at limiting $C_i$ conditions (Van and Spalding, 1999). An explanation for these conflicting results may lie in the fact that the physiology was tested at near pH 7 where $HCO_3^-$ and $CO_2$ are almost equal in abundance. It has been suggested that the defective phenotype may be only apparent at high pHs or that an increased contribution from the Cah2 may alleviate potential limitations (Moroney *et al*., 2001). This high pH condition can be found where unstirred boundary layer conditions are high such as in algal films rather than in dilute

liquid culture. Cah2 is not expressed under low $CO_2$ conditions and is never expressed at high levels. The function of Cah2 is not clear.

Other proteins upregulated under low $CO_2$ condition include a 53 kDa alanine aminotransferase (Chen *et al*., 1996) and two 21 kDa mitochondrial β-CAs (Karlsson *et al*., 1995). During photorespiration $CO_2$ is produced. $CO_2$ inhibits the mitochondrial respiration (Gonzàlez-Meler *et al*., 1996) and this mitochondrial CA could be involved in the efflux of this $CO_2$. This CA can stabilize the pH in the mitochondria during photorespiration by increasing the rate at which photorespiratory $NH_3$ is converted to $NH_4^+$ with uptake of one $H^+$ from the mitochondrial matrix (Eriksson *et al*., 1998). It has been suggested that these mitochondrial CAs play a role in recycling both respiratory and photorespiratory $CO_2$ by converting it to $HCO_3^-$ in the mitochondrial matrix (Raven, 2001). The $HCO_3^-$ would leak back into the cytosol where it would be available for transport into the chloroplast. This model assumes the absence of CA from the cytosol. Recently, it has been shown that the expressions of these mitochondrial CAs decrease when the external $NH_4^+$ concentration decreases, to the point of being undetectable when the $NH_4^+$ supply restricts the rate of photoautotrophic growth (Giordano *et al*., 2003). The expression of these CAs were induced at 0.2% $CO_2$ condition by increasing the $NH_4^+$ condition in the growth medium. The workers have proposed that the mitochondrial CAs are involved in supplying $HCO_3^-$ for anaplerotic assimilation catalyzed by phosphoenolpyruvate carboxylase, which provides carbon skeletons for nitrogen assimilation under certain conditions.

Two other polypeptides (45 kDa and 42 kDa) are expressed under low $CO_2$ conditions and their identities are still not clear. Absence of any of these proteins may be correlated to an inability for growth under low $CO_2$ conditions and indicate that these proteins play a role in CCM. The 36 kDa protein (LIP-36) identified in the labeling studies and subsequently purified (Spalding and Jeffrey, 1989; Geraghty *et al*., 1990) has been localized specifically to chloroplast

envelope membranes (Mason *et al*., 1990; Ramazanov *et al*., 1993). It has been shown to be encoded by two genes and the protein sequence shows homology to a mitochondrial carrier protein superfamily (Chen *et al*., 1996).

Analysis of the CCM mutant *pmp1*, shows that it synthesizes all the known low $CO_2$ inducible polypeptides under low $CO_2$ conditions except the 42 and 45 kDa proteins. This mutant accumulates $C_i$ but to levels lower than the wild type. Another mutant, *cia5,* has a HCR phenotype, fails to induce CCM and does not synthesize any of the low $CO_2$ inducible polypeptides. It is deficient in a putative transcription factor (Fukuzawa *et al*., 2001; Xiang, *et al*, 2001).

**d. Pyrenoid, Rubisco Localization and Its Role in CCM**

The pyrenoid, thought to be the structural equivalent of the carboxysome, is found in most eukaryotic green algae. Pyrenoids have been purified from both *Eremosphera* (Okada, 1992) and *C. reinhardtii* (Kuchitsu *et al*., 1991) and in both cases they consisted primarily of Rubisco. In addition, *C. reinhardtii* cells with a mutation in the rbcL gene (Rubisco large subunit) that leads to a truncation of the large subunit of Rubisco have no pyrenoids (Rawat *et al*., 1996). Although it is accepted that Rubisco is the major constituent of the pyrenoid, there are conflicting findings regarding what percentage of the cell's Rubisco is in the pyrenoid. Cryofixation and immunogold studies have indicated that the pyrenoid in the low $CO_2$ grown cells contain almost 90% of the Rubisco (McKay and Gibbs, 1991; McKay *et al*., 1991; Morita *et al*., 1997; Borkhsenious *et al*., 1998). *In vitro* measurements of Rubisco activity imply that the enzyme in the pyrenoid must be active to account for the $CO_2$ fixation rate observed in *C. reinhardtii*. Under high $CO_2$ conditions only 50% of the Rubisco is found in the pyrenoid (Borkhsenious *et al*., 1998). In lichens and bryophytes there is a good correlation between the operation of a CCM and the presence of a pyrenoid (Smith and Griffiths, 1996). A specific

location of Rubisco is also compatible with the view that organisms that have CCMs specifically package Rubisco. In *C. reinhardtii*, a rapid formation of starch sheath has been observed in response to low $CO_2$. This pyrenoid starch sheath formation is partially inhibited by the presence of acetate in the growth medium under light and low $CO_2$ conditions. When cells are grown in the dark, the CCM is not induced and the starch sheath is not formed even though there is a large amount of starch accumulation in the chloroplast stroma. The membrane permeant CA inhibitor EZ inhibits formation of the starch sheath under low $CO_2$ conditions (Ramazanov *et al*., 1996). Some of the HCR mutants partially or fully lack the starch sheath. These observations suggest that the ultrastructural reorganization of the pyrenoid starch sheath under low $CO_2$ conditions may play a role similar to the carboxysome in cyanobacteria (Borkhsenious *et al*., 1998).

The thylakoid α-CA (Cah3) appears to be targeted to the thylakoid lumen and is associated with PSII (Karlsson *et al*., 1998; Villarejo *et al*., 2002; Park *et al*., 1999). A *C. reinhardtii* mutant *cia3*, that has two point mutations in the region coding for the transit peptide of the Cah3 protein, grows poorly under low $CO_2$ but grows fine under high $CO_2$ (Fig. 1.6). *ca-1*, another allelic mutant of *Cah3* showed the same phenotype under low $CO_2$. The sequence of the *ca-1* gene revealed a point mutation in the 5′ end of the gene that created a stop codon. *cia3* and *ca-1 allelic* mutants were complemented by the wild type *Cah3* gene (Karlsson *et al*.,1998; Funke *et al*., 1997) and the transformants resembled wild type cells and grew fine under low $CO_2$. This shows that Cah3 is required for the growth of *Chlamydomonas* in air levels of $CO_2$ (Funke *et al*., 1997).

It had been initially suggested that this CA speeds up the formation of $CO_2$ from $HCO_3^-$ in the acidic lumen of thylakoids and that this $CO_2$ diffuses through the thylakoid membrane to the pyrenoid (Moroney and Mason, 1991; Badger and Price, 1994; Raven, 1997; Moroney and Somanchi, 1999). This model is based on the assumption that $HCO_3^-$ is actively pumped into the

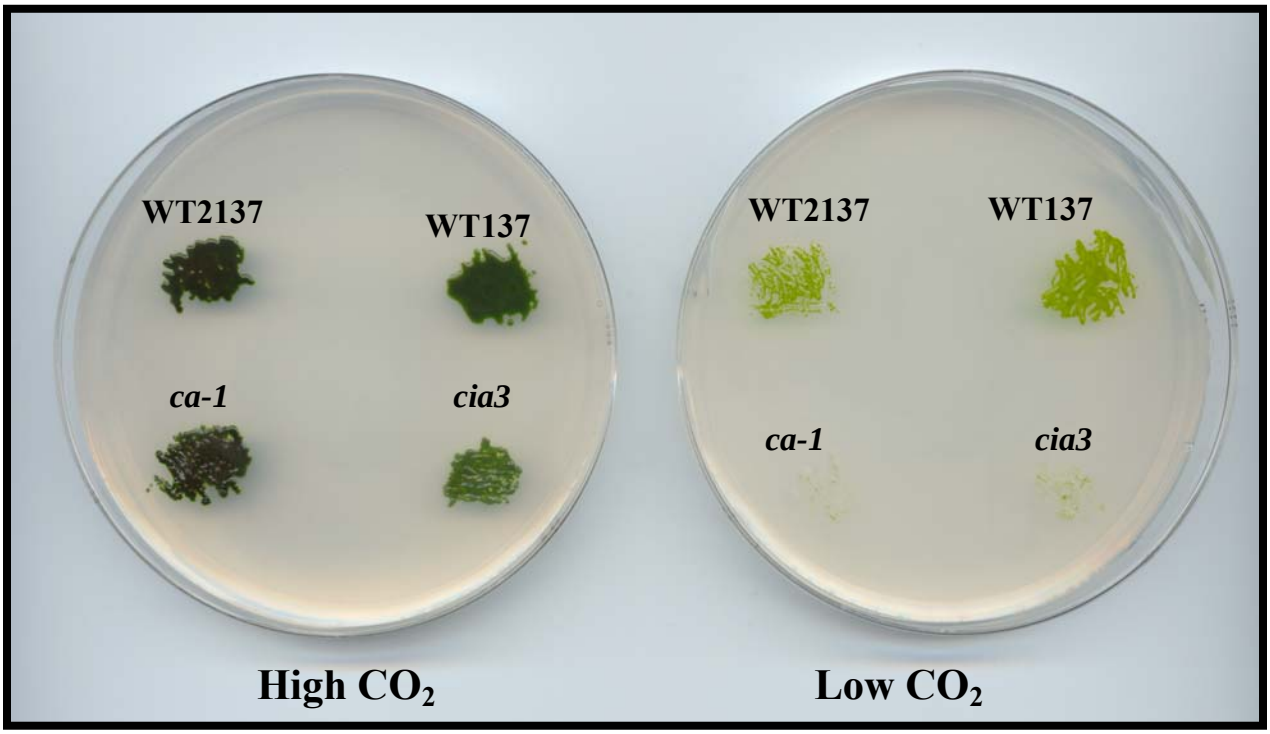



**Figure 1.6 Growth of the wild type and the *cah3* mutant cells of *C. reinhardtii* on the minimal media under low and high $CO_2$ conditions.** *ca-1* and *cia3* are allelic *cah3* mutants. WT2137 is the parent of *ca-1* mutant strain and WT137 is the parent of *cia3* mutant strain.

lumen from the stroma and also on the fact that in many algae, including *Chlamydomonas*, thylakoids penetrate the pyrenoid matrix in various patterns (Raven, 1997).

It has also been proposed that the activity of this CA stabilizes the manganese cluster of PSII and maintains the oxygen evolution complex (OEC) in a functionally active state (Villarejo *et al*., 2002). In the mutant *cia3*, the OEC contains two manganese atoms per PSII reaction center and shows very low OEC activity. Addition of $HCO_3^-$ or manganese to the thylakoids and BBY (Berthold-Babcock-Yocum PSII particles) preparations from the mutant can restore the wild type OEC activity (Villarejo *et al*., 2002). Recently, it has been proposed that the thylakoid CA is mainly required for the proper functioning of CCM under low $CO_2$ conditions by providing elevated levels of $CO_2$ to Rubisco. It might have the secondary role of stabilizing PSII at high light, possibly by preventing an excessively low luminal pH through rapid dehydration of $HCO_3^-$ to $CO_2$ and $H_2O$ (Hanson *et al*., 2003). To date, there is no report of a thylakoid localized CA in higher plants. This also seems to implicate that probably Cah3 is not essential for the proper functioning of PSII.

There may be diversity in the ways in which $CO_2$ is generated in the chloroplast and where it is elevated (Raven, 1997; Badger *et al*., 1998). For example both the pyrenoid and thylakoid lumen based CAs might release $CO_2$ from $HCO_3^-$ pool in the stroma. There may also be a chloroplast envelope CA. The possibility that $CO_2$ may be elevated within the entire chloroplast rather than just the pyrenoid has also been suggested (Badger, 2003).

Thylakoid associated α-CA has been found also in two other green microalgae, *Tetraedron minimum* and *Chlamydomonas noctigama* (van Hunnik *et al*., 2001). Some arctic species of unicellular green algae such as *Chloromonas* and the lichen photobionts like *Coccomyxa* lack both pyrenoids and the CCM (Palmqvist *et al*., 1994; Morita *et al*., 1997). Therefore it is speculated these algae lacking both pyrenoids and CCM may be slow growing

organisms adapted to environments in which CCM might not be necessary for survival. (Honegger, 1991).

Recent data on the physiology of carbon fixation in the marine diatom *Thalassiosira weissflogii* suggest that a form of $C_4$ photosynthesis may operate together with active $HCO_3^-$ uptake (Reinfelder *et al*., 2000). In this respect a novel δ-CA class of enzyme has been identified which is unrelated to other three classes of CAs (Cox *et al*., 2000). This enzyme appears to be localized in the cytosol and presumably could participate in providing $HCO_3^-$ to PEP carboxylase (Morel *et al*., 2002). However, there is some controversy regarding the data and further investigation will be needed to resolve the picture. Recently a β-CA has been identified in another marine diatom *Phaeodactylum tricornutum* (Satoh *et al*., 2001) although its localization is uncertain.

**5. CCM in Bryophytes**

Bryophytes are among the first green plants to colonize the terrestrial environment. They are closely related to the higher plants that dominate the land today (Nickrent *et al*., 2000). There is less discrimination against $^{13}CO_2$ during photosynthetic carbon fixation in bryophytes. This indicates that a carbon concentrating mechanism is present in these plants. Anthocerotophyta (hornworts), a group of Bryophytes, have unique carbon concentrating properties. Many members of this group have one chloroplast per cell and a pyrenoid body is present in the chloroplast. These organisms have increased affinities for $CO_2$, reduced $CO_2$ compensation points, accumulate $C_i$ and to some extent show reduced $O_2$ sensitivity of photosynthesis (Smith and Griffiths, 1996; Hanson *et al*., 2002). The CCM in these plants is hampered when they are treated by the CA inhibitor EZ (Smith and Griffiths, 2000). This indicates a role for CA similar to other algal chloroplast based CCMs. No information about CAs in these organisms is known.

## OTHER FUNCTIONAL ROLES PLAYED BY CAs IN DIFFERENT GROUPS OF ORGANISMS

Carbonic anhydrase has been implicated in many physiological processes such as carboxylation/decarboxylation reactions, transport of $CO_2$ and/or $HCO_3^-$ across membranes either by passive or active mechanisms, pH regulation, ion exchange, calcification, metabolism of urea, glucose and lipids and efficient viral replication. Membrane associated human CAXVI appears to be associated with plasma membrane of cells of the proximal convoluted tubule of the kidney suggesting a role in the reabsorption of $HCO_3^-$ by that organ. Two recently characterized transmembrane proteins, CAIX and CAXII, have been linked to oncogenesis and their overexpression has been observed in malignant tumors (Türeci *et al*., 1998; Ivanov *et al*., 1998). Recently a RNA and DNA binding nuclear factor, nonO/p54$^{nrb}$ has been shown to possess CA activity (Karhumaa *et al*., 2000). The activity of nonclassical CA nonO, is higher than CAIII and CAV and may function in the maintenance of pH homeostasis in the nucleus.

Deletion of the β-CA-like gene *NCE103* of the yeast *Saccharomyces cerevisiae* causes an oxygen- sensitive growth defect (Gőtz *et al*., 1999). The protein has no detectable CA activity. Expression of the *Medicago sativa* CA gene in a yeast expression cassette on a multicopy plasmid complemented the growth defects caused by the deletion of the *NCE103* gene. This shows that the *Nce103* protein is required for protection against oxidative metabolism. Recently it has been found that the tobacco salicylic acid-binding protein 3 (SABP3) is a chloroplast β-CA, which exhibits antioxidant activity and plays a role in the hypersensitive defense response (Slaymaker *et al*., 2001).

In *Escherichia coli*, a eubacterium, cyanate induces the expression of *cyn* operon. The *cyn* operon includes the gene *cynS* which encodes a cyanase that catalyzes the reaction of cyanate with $HCO_3^-$ to give ammonia and $CO_2$. The *cynT* gene which codes for a β-CA, is a part of the

*cyn* operon. This β-CA recycles the $CO_2$ produced in the cynase reaction back to $HCO_3^-$ which would have diffused out of the cell (Guilloton *et al*., 1993).

*Methanosarcina thermophila* is a methanogen in that it obtains energy for growth by converting the methyl groups of acetate, methanol or methanolamines to methane. During metabolism of acetate, oxidation of the carbonyl group provides electrons for reduction of the methyl group leaving $CO_2$ as the second product. A metabolic switch from methanol to acetate elevates γ-CA activity suggesting this enzyme is important for growth on acetate. It has been proposed that CA may be required for a $CH_3CO_2^-/H^+$ symport system or for efficient removal of cytoplasmically produced $CO_2$ (Alber and Ferry, 1994).

An α-CA has been recently identified throughout young nodules in soybean and mainly in the cortical regions of old nodules (Kavroulakis *et al*., 2000). This suggests that this CA recycles $CO_2$ early in the nodule development and facilitates diffusion of $CO_2$ from the nodule later in development.

There has been an increasing interest in CAs from plants and algae over the past decade. This interest began with the discovery of the β-CA in plants in 1990 (Fawcett *et al*., 1990) and has continued with the finding of multiple α- and β-CAs in *C. reinhardtii* and *A. thaliana* and the determination of the critical physiological roles CAs have in cyanobacteria and macro-algae. There are still numerous unresolved questions regarding the number and distribution of CA genes and gene families, structure, localization and function. The availability of *Arabidopsis* and *Chlamydomonas* genome sequences can be used to find out the exact number of CA isoforms in these organisms. The challenge for future researchers will be to determine the expression patterns, localization and physiological roles for each of these isoforms. As there appears to be a large number of isoforms in plants and algae, CA researchers have a lot more to work on in the near future.

This dissertation is the report on my efforts to a) identify, clone, and overexpress novel CA and CA-like genes from the green unicellular algae *Chlamydomonas reinhardtii* in *Escherichia coli*, b) partially biochemically characterize the recombinant CA protein if it is active, and c) determine if the new active CA plays any role in photosynthesis and CCM in this green alga.

# CHAPTER 2

# MATERIALS AND METHODS

## CELL CULTURE

Wild type *C. reinhardtii* $137^+$ and $2137^+$ were obtained from Dr. R. K. Togasaki, of Indiana University, Bloomington, and from *Chlamydomonas* Genetic Center at Duke University, Durham, respectively. The strain D66 ($nit2^-$, cw15, $mt^+$) obtained from Rogene Schnell, University of Arkansas-Little Rock (Schnell and LeFebvre, 1993) and *ca-1* was obtained from Dr. M. Spalding, of Iowa State University, Ames, Iowa. Strains c*ia3* and the *cia3* transformant 9A were generated by Dr. J. V. Moroney's laboratory group (Moroney *et al*., 1986; Karlsson *et al*., 1998). To start cultures, cells from yeast acetate medium plates were inoculated into 100 mL of TAP (Tris-Acetate-Phosphate) medium (Sueoka, 1960) and grown with continuous shaking and light (300 μmol photons • $m^{-2}$ • $s^{-1}$) for two days. An aliquot of the culture was then transferred to 1.5 L of minimal medium (Sueoka, 1960) and bubbled with high $CO_2$ (5% $CO_2$ in air) until it reached a cell density of about 2 x $10^6$ cells $mL^{-1}$. The culture was diluted with an equal volume of fresh medium and split into two flasks. One was bubbled with high $CO_2$ and the other with low $CO_2$ (0.035% $CO_2$ in air). The time of low $CO_2$ acclimation varied from 2 to 12 hours. The high and low $CO_2$ acclimated D66 cells were used for RNA isolation, measurement of chlorophyll content and Western blots. Only low $CO_2$ acclimated cells of strains D66, $137^+$, *ca-1*, *cia3* and 9A were used for chlorophyll assays and Western blots.

## TOTAL RNA ISOLATION AND NORTHERN BLOT ANALYSES

All glassware used for RNA isolation was either heated to 180°C overnight or soaked with DEPC (diethyl pyrocarbonate)-treated water and autoclaved for two hours. Total RNA was

extracted from *C. reinhardtii* using standard procedures (Sambrook *et al*., 1989). Two liters of cultures of either high or low $CO_2$ grown cells were harvested and resuspended in 10 mL of DEPC-treated water in a 150 mL Corex tube. To this, 20 mL of 2X lysis buffer [100 mM Tris-HCl, 4% SDS, 3 M NaCl and 30 mM EDTA (ethylenediaminetetraacetic acid; pH 8); made with DEPC-treated water and autoclaved for one hour] and 3 mg of proteinase K (10 mg $mL^{-1}$) were added. The final concentration of proteinase K was 100 μg $mL^{-1}$. The mixture was gently stirred at room temperature for 20 minutes. The cell lysate was extracted three times with phenol/chloroform/isoamyl alcohol (25:24:1,v/v) and once with chloroform. After centrifugation at 2500 rpm in a swinging bucket rotor for 15 minutes, the supernatant was transferred to a clean tube to which an equal volume of isopropanol was added. The tube was kept at -20°C overnight and the RNA was collected by centrifugation at 15000 g for 20 minutes at 4°C.

The RNA pellet was washed once with 80% ethanol and air dried for several hours. The RNA was dissolved in 4 mL of DEPC-treated water. LiCl (12 M) was added to the solution to achieve a final concentration of 2.5 M and the RNA was precipitated overnight at 4°C. RNA was collected by centrifuging at 15000 g for 20 minutes at 4°C. After being washed, the RNA pellet was dried and resuspended in 6 mL of DEPC-treated water. RNA was reprecipitated by adding 2.5 volumes of ethanol at -20°C for 4 hours. The recovered RNA was washed, air dried and resuspended in 1 mL of DEPC-treated water.

All RNA samples were aliquoted (10 μg in each tube) and stored at -70°C. Total RNA concentration was determined by absorbance at 260 nm, using the conversion 1 $OD_{260}$ = 40 μg of RNA $mL^{-1}$. For Northern blot analyses, 5.5 μL of RNA (20 μg) was added to 1 μL of 10X MOPS (morpholinepropanesulfonic acid) buffer, 3.5 μL of formaldehyde, 10 μL of formamide and incubated at 65°C for 15 minutes. Four μL of loading buffer [50% glycerol, 1 mM EDTA (pH 8), 0.25% bromophenol blue, 0.25% xylene cyanol FF] was added after the samples were

chilled on ice for 3 minutes. The denatured RNA extracted from high or low $CO_2$ adapted cells was resolved on a 1% agarose gel containing 6% formaldehyde [gel running buffer: 1 X MOPS, 6% formaldehyde; 10 X MOPS gel running buffer: 0.2 M MOPS (pH 7.0), 80 mM sodium acetate, 10 mM EDTA; the solution was filter sterilized or autoclaved for 15 minutes]. The gel was soaked overnight in DEPC-treated water and transferred to a BA-S 85 nitrocellulose membrane (Schleicher and Schuell Bioscience Inc, Keene, New Hampshire). The Northern blot was probed with the radiolabeled *Cah6* PCR product described in the next section. The final washing conditions were 0.5 X SSC + 0.1% SDS at 55°C for 2 hours or longer.

**RANDOM LABELING OF DNA TO PRODUCE THE RADIOACTIVE PROBE**

When *Cah6* primers X-9 and R5 were used to perform PCR on the cDNA core library, an 826 bp PCR product is generated. This product was used to make a radioactive probe to study the expression pattern of *Cah6* under low and high $CO_2$ conditions. The PCR product corresponded to an 826 bp region in the 3′ UTR of *Cah6*. This region was selected as it was unique to *Cah6* and not similar to any sequence in the two mitochondrial β-CAs, present in *C. reinhardtii*. Sequences of the X-9 and R5 primers are given in Appendix 1. One µL of 50 ng $\mu L^{-1}$ random hexanucleotide primers was mixed with the 50 ng of purified *Cah6* PCR product in a 0.5 mL centrifuge tube and the final volume was adjusted with water to 10 µL. After the DNA and the primer mixture were boiled for 5 minutes and instantly chilled on ice for 10 minutes, the following components were added to the tube: 2 µL of 10 X DNA I buffer (NEB, Beverly, Massachussets), 2 µL of dNTP (2.5 mM each; no dCTP), 5 µL of $^{32}$P-dCTP (10 µCi $\mu L^{-1}$). The tube was briefly centrifuged and 1 µL of DNA polymerase I large fragment (Klenow) was added and mixed gently. The reaction was incubated at room temperature for 4 hours. The probe was purified by passing the mixture through G-50 spin column (1000 g x 3 minutes). The *Cah6* probe had a specific activity greater than 2 x $10^8$ cpm $\mu g^{-1}$ DNA.

**DNA PREPARATION, SEQUENCING AND HOMOLOGY ANALYSES**

Total DNA was isolated from wild type D66 cells grown as patches on TAP plates according to Newman *et al*. (1990). Briefly, the cells were resuspended in disruption buffer containing SDS and then the nucleic acids were extracted using phenol/chloroform. The aqueous phase was extracted a second time with choloform. The nucleic acids were then precipitated with an equal volume of ethanol and washed twice with 70% ethanol.

Plasmid and cosmid DNA was purified using a combination of the standard ethanol precipitation method (Sambrook *et al*., 1989) followed by the purification method using the spin columns from a commercial kit (QIAGEN, Chatsworth, California). cDNA and genomic PCR products were purified from the 0.8% agarose gels. Amplified bands were cut from the gel and were treated with 6 M NaI at 55°C to melt the gel piece. DNA was purified from the liquefied gel using the mini spin columns from the commercial kit mentioned above. DNA molecular size standards were 2-log and 1 kb DNA (NEB, Beverly, Massachussets). For DNA concentration determination, HindIII digest of λ DNA was used as a standard (NEB, Beverly, Massachussets).

DNA was sequenced using ABI dye terminators (Perkin Elmer Applied Biosystems, Foster City, California); for some PCR fragments and cosmids enriched in the GC content, the use of dGTP-BigDye generated better sequences than dITP-BigDye). For plasmid clones greater than 6 kb, 800 ng of DNA was used per sequencing reaction. One-two µg of DNA was used to sequence cosmids. Ten and 40 ng of DNA were used to sequence PCR products (100-500 bp) and PCR products (500-1000 bp), respectively. The sequencing reaction contained 3.2 pMol of primers and 2 µL of Big Dye. The following PCR profile was used: 30 cycles at 96°C for 30 seconds, 50°C for 15 seconds and 60°C for 4 minutes. Homology searches (against *Chlamydomonas* EST and the full database) were performed using the BLAST server (http://www.ncbi.nlm.nih.gov/BLAST) (Altschul *et al*., 1997). Exon/intron splice sites and open

reading frames were identified manually (Silflow, 1998) as well as by using GreenGenie (Kulp *et al*., 1996). Signal peptide analysis and molecular weight and pI calculations were determined with different protein prediction programs like CHLOR P, TARGET P and SORT P which had hyperlinks in the ExPasy server (http://ca.expasy.org/tools/#translate).

**cDNA LIBRARY PREPARATION AND AMPLIFICATION OF *CAH6* AND *GCLP1* cDNA FROM THE PREPARED LIBRARY**

A cDNA core library (amplified bacteriophage libraries) was obtained from the *Chlamydomonas* Genetics Center at Duke University, Durham. This core library was made from cDNAs prepared from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of $CO_2$ (0.035% $CO_2$) and HS medium bubbled with 5% $CO_2$. cDNAs were cloned into the lambda Zap II (Stratagene, La Jolla, California) in the EcoRI (5') and XhoRI (3') sites. The lambda ZAP II vector is designed to allow simple, efficient, *in vivo* excision and recircularization of any cloned insert contained within the lambda vector to form a phagemid containing the cloned insert. *In vivo* excision, involving the ExAssist interference-resistant helper phage along with the SOLR strain of *E. coli,* was used. The ExAssist helper phage contains an amber mutation that prevents the replication of the phage genome in a nonsuppressing *E. coli* strain like SOLR. This allows efficient excision of the pBluescript phagemid from the Lambda ZAP II vector while eliminating the problems associated with helper phage co-infection.

Overnight cultures (50 mL) of XL1 Blue MRF´ or SOLR cells were grown in LB broth at 30°C. Harvested cells were resuspended in 25 mL of 10 mM $MgSO_4$. The concentration of cells in 10 mM $MgSO_4$ was adjusted to a concentration of 1 x $10^8$ cells $mL^{-1}$. The amplified lambda bacteriophage library was mixed with the XL1 Blue cells in the ratio of 1:10. This was done to excise 10 to 100 fold more lambda phage than the size of the primary library to ensure statistical

representation of the excised clone. ExAssist helper phage was added at a 10:1 helper phage to cells ratio to ensure that every cell was co-infected with the lambda phage and the helper phage. The mixture was allowed to incubate at 37°C for 15 minutes to allow the phage to attach to the cells. Twenty mL of fresh LB broth were added to the mixture and allowed to incubate for 2-3.5 hours at 37°C with shaking. The mixture was then heated to lyse the phage particles and the cells. The lysed mixture was spun at 1000 g for 10 minutes to pellet the cell debris and the supernatant containing the excised phagemid was collected. The titer of the excised phagemid in the supernatant was determined and was used to calculate the number of colonies that were needed for the statistical representation of the excised clones. Colonies from all the culture plates were pooled by mass scraping and resuspension in 3 mL of LB. Plasmids were purified from these cells using the spin column method (QIAGEN, Chatsworth, California).

*Cah6* primers F4 and R5 (Appendix 1) when used on the cDNA core library yielded a PCR product of 2452 bp. This product codes for the full length Cah6 protein. *Gclp1* primers M1F and M1R yielded a PCR product of 1170 bp which codes for the full length Gclp1 protein.

**SCREENING OF THE COSMID LIBRARY**

To obtain wild-type genomic clones of the *Cah6* and *Gclp1* genes, a PCR-based screen of an indexed cosmid library was used. An indexed cosmid library was constructed using a cosmid library from Saul Purton, University of London (Purton and Rochaix, 1994). Briefly, 7680 different *E. coli* lines carrying single cosmids were grown in LB media on 80 different 96-well microtiter plates. Using this indexed library, 80 pools of cells, each containing 96 single cosmids, were generated. DNA from each pool, obtained by common alkaline lysis procedures (Sambrook *et al*., 1989), was used to create 10 superpools (each containing about 768 single cosmids) that were suitable for PCR. Using the sequence information obtained from the EST database of *C. reinhardtii*, primer sets were designed and used to screen the superpools. Cah6

primers F4 and R4 (Appendix 1) were used for PCR. These primers amplified a *Cah6* product of 1.8 kb. Once a plate carrying the correct cosmid was identified, a new set of pools was generated (12 pools, each containing 8 single cosmids). Finally, a new PCR reaction was performed with the individual cosmids from the positive pool described above. Using this protocol, after 4 rounds of PCR, two cosmids containing the *Cah6* gene, designated 72-E-6 and 29-D-12, were isolated from the cosmid library. PCR using primers F4 and R5 yielded a product of 2879 bp when used on the isolated cosmid clones. This PCR product is just six base pairs short from being the full length genomic sequence of *Cah6*.

The *Gclp1* PCR primers 1F and 2B (Appendix 3) were used on genomic DNA to generate a product of 745 bp. This PCR product was used to make a probe to screen the indexed cosmid library. After three sequential rounds of screening, two cosmid clones namely 70-C-3 and 70-G-8 were isolated. Primers M1F and M1R were used on the two isolated cosmid clones to yield a PCR product of 2249 bp. The sequences of the *Cah6* and *Gclp1* primers and their alignments on the corresponding genomic maps are given in Appendix 1 and Appendix 3, respectively.

**PRODUCTION OF OVEREXPRESSION CONSTRUCTS**

*Cah3*, *Cah6* and *Gclp1* were cloned into the pMal-c2x overexpression vector (NEB, Beverly, MA) downstream from the *MalE* gene which encodes maltose-binding protein (MBP). The vector (6648 bp) has an exact deletion of the *MalE* signal sequence (bases 1531-1605) resulting in the cytoplasmic expression of the fusion protein. The vector contains the inducible Ptac promoter, which is a hybrid of Trp promoter and LacUV5 promoter. Ptac is positioned to transcribe a *MalE-LacZα* gene fusion (Fig 2.1). The $LacI^q$ gene codes for the Lac repressor and has a promoter mutation which increases intracellular concentration of LacI repressor. It turns off transcription from Ptac until IPTG (Isopropyl-β, D-thiogalactopyranoside) is added (Fig. 2.1).

The polylinker cloning region provides restriction endonuclease sites to insert the gene of interest, fusing it to the *MalE* gene (Fig 2.2). Insertion of the desired gene into this site interrupts the *LacZα* ORF (open reading frame) allowing a blue-white selection on LB + Amp + X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) + IPTG plates. The vector also has *Amp*$^r$ gene which codes for the β-lactamase gene that confers ampicillin resistance. This allows additional screening of transformants on LB + Amp plates. The fusion protein can be purified by one-step affinity chromatography using an amylose column. The vector has a spacer sequence coding for 10 aspargine residues between the *MalE* sequence and the polylinker sequence (Fig. 2.2). This allows insulation of MBP from the protein of interest, increasing the chances that a particular fusion protein will bind tightly to the amylose resin during purification. The vector also includes a sequence coding for the recognition site of a protease, Factor Xa, which allows the desired protein to be cleaved from MBP after purification.

The *Cah3* cDNA sequence coding for the mature Cah3 protein was cloned into the pMal-c2x overexpression vector. "A" and "B" are the 5′ and 3′ end PCR primers, respectively used for amplification of *Cah3* cDNA (790 bp). The primer A has the sequence 5′- AAG GGA TCC ATC GAA GGG CGC GCA GCT TGG ACC TAT TAT GGC GAA -3′ and the primer B has the sequence 5'- GTA AAG CTT CCC CAC CGT GGG CCA AAC -3'. The primer A has a Bam HI and a Factor Xa site incorporated at the 5′ end. Primer B has a HindIII site incorporated at the 5′end. The vector was double digested with HindIII and BamHI and ligated to the 790 bp Cah3 PCR product. This recombinant construct had two different Factor Xa cutting sites and was provided by Dr. Göran Samuelsson, Umeå University, Sweden. Two different pMal-*Cah6* recombinant expression constructs were designed. One construct contained the cDNA sequence coding for the full length Cah6 protein and the other contained the cDNA sequence that would code for the putative mature Cah6 protein. SP2 + R4H primers and MP + R4H primers were

# RECOMBINANT pMAL CONSTRUCT

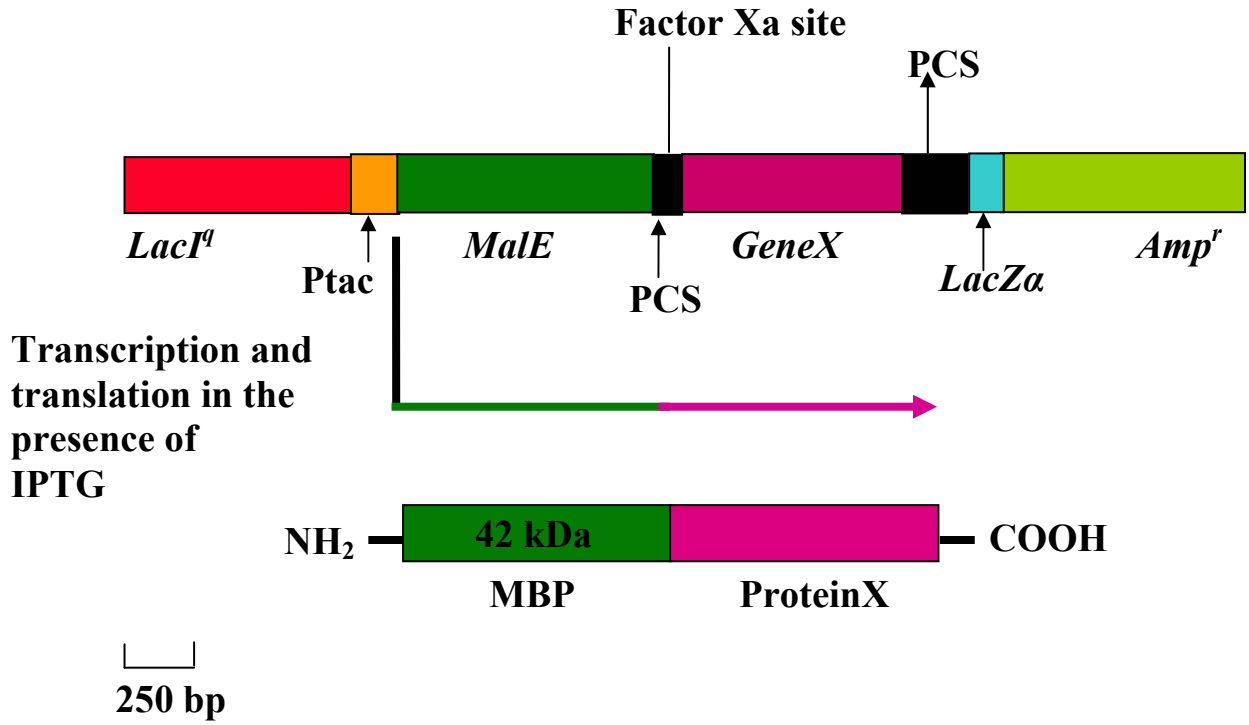



**Figure 2.1 A diagram showing some features of the pMal expression vector and the translation of the recombinant construct.** *LacI*$^q$ stands for an allele that has a promoter mutation which increases intracellular concentration of LacI repressor, Ptac is a hybrid of Trp promoter and LacUV5 promoter, PCS stands for the polylinker cloning site, *Gene X* is any gene cloned into the vector, *LacZα* stands for the β-galactosidase gene and *Amp*$^r$ stands for the β-lactamase gene.

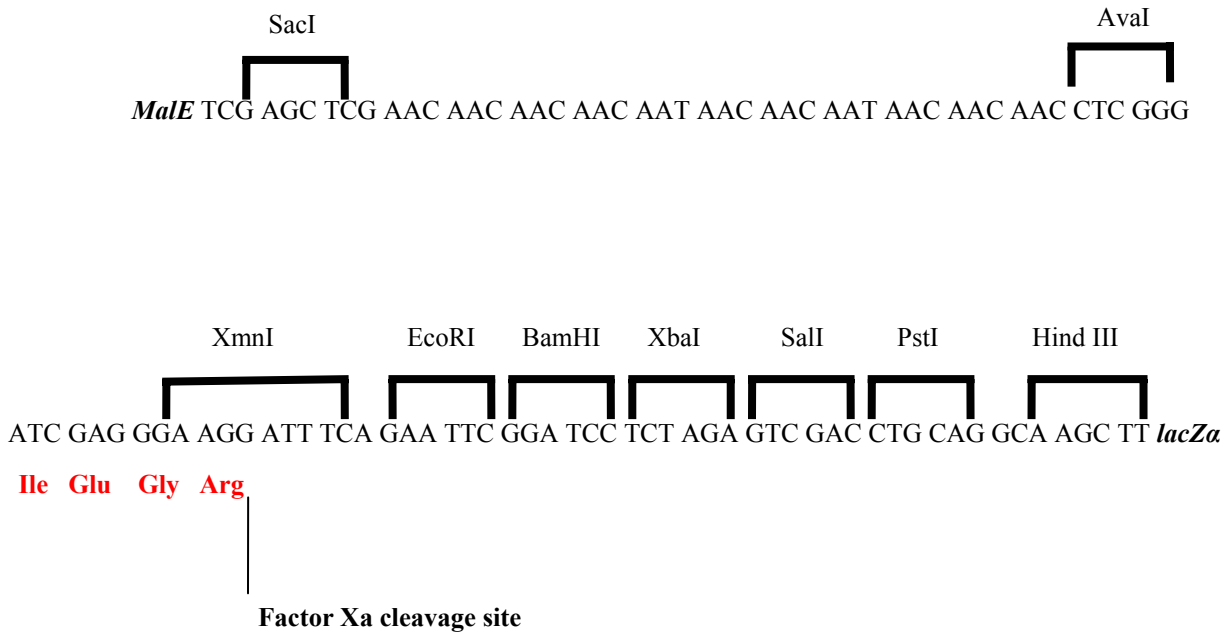



**Figure 2.2 Partial sequence of the polylinker cloning site of pMal-c2x.** The Factor Xa recognition amino acid sequence is shown in bold red. The Factor Xa cleavage site is denoted by a black line. *MalE* codes for the maltose binding protein (MBP) and *LacZα* codes for the β-galactosidase α-fragment. Restriction enzyme sites in the polylinker cloning region are shown on top.

used to amplify the cDNA coding for the full length and mature Cah6 protein, respectively. SP2 + R4H and MP + R4H primers yielded a PCR product of 1.1kb and 979 bp, respectively.

Like MBP-Cah6 constructs, two different MBP-Gclp1 expression constructs were made. S1 + M1RH primers and S3 + MIRH primers (Appendix 3) were used to amplify the cDNA coding for the full length and putative mature Gclp1 protein, respectively. S1 + M1RH and S3 + M1RH primers amplified a 1067 bp and 944 bp product, respectively.

A high fidelity DNA polymerase (Platinum Pfx DNA polymerase from Invitrogen, Carlsbad, California) was used for all PCRs. This was used to prevent any missense/null mutation in the protein sequence during PCR and generated blunt ended amplified products. The vector was double digested with the XmnI and HindIII. Amplified cDNAs were purified from the DNA gel using QIAGEN spin columns and were digested with HindIII. Ligation of the insert to the overexpression vector pMal-c2x vector was performed following the protocol in the New England BioLab technical manual.

Transformations of DH5α cells were performed following the protocol in Sambrook *et al*. (1989). Although transformants were isolated by screening both on LB + Amp (100 μg $mL^{-1}$) and LB + Amp (100 μg $mL^{-1}$) + IPTG (Isopropyl-β, D-thiogalactopyranoside; 1 mM) + X-Gal (80 μg $mL^{-1}$) plates, they were picked only from LB + Amp plates. Because of the strength of the Ptac promoter, transformants taken from a plate containing IPTG can contain mutant plasmids that have either, 1) lost part or all of the fusion gene, or 2) no longer express the protein at high levels.

Ampicillin (50 mg/mL) and IPTG (200 mM) stocks used for the experiment were filter sterilized. In-frame insertion of *Cah6 and Gclp1* with the sequence of MBP in the recombinant clone was verified by double restriction enzyme digestion analyses with SacI and HindIII and DNA sequencing.

**OVEREXPRESSION AND PURIFICATION OF MBP FUSION PROTEINS**

Selected clones of *Cah3*, *Cah6* and *Gclp1* were grown at 37°C in 2 L LB + glucose (0.2%) + Amp (100 μg $mL^{-1}$) cultures on a rotary shaker. Glucose is necessary in the growth medium to repress the maltose genes on the chromosome of the *E. coli* host, one of which codes for amylase which can degrade the amylose on the affinity resin that is used for purification. The cells were induced for 2 hours with 1 mM IPTG at 37°C when the culture $OD_{600}$ was between 0.6 - 0.7. Both induced and uninduced cells were harvested and resuspended in column buffer [20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA (ethylenediaminetetraacetic acid) with or without 10 mM 2-mercaptoethanol] and ruptured in a prechilled French pressure cell. Equal amounts of protein samples of ruptured induced and uninduced cells were loaded on a 12% SDS-polyacrylamide gel and subjected to electrophoresis to verify the overexpression of the recombinant protein.

The fusion protein was purified by one-step affinity chromatography using amylose resin (Fig. 2.3). Amylose resin (1 mL of amylose resin binds 3 mg of the recombinant protein) was mixed with the crude ruptured cell extract on a shaker at room temperature for three hours at room temperature and poured into a 2.5 cm x 10 cm column to perform batch purification. The column was washed with 6 L of column buffer to remove other proteins. At the final step, fusion proteins were eluted from the column by column buffer containing 10 mM maltose. Purified recombinant fusion proteins were further concentrated by a passage through the 100 kDa centricon columns (Amicon, Billerica, Massachussets). The recombinant proteins were recovered from the membrane of the filter in the centricon columns.

Recombinant proteins were cleaved from the maltose binding protein (MBP) by digestion with the protease Factor Xa (NEB). Fifty μg of the recombinant protein were digested by 1 μg of Factor Xa enzyme in the Factor Xa digestion buffer [20 mM Tris-HCl, 100 mM NaCl, and 2 mM

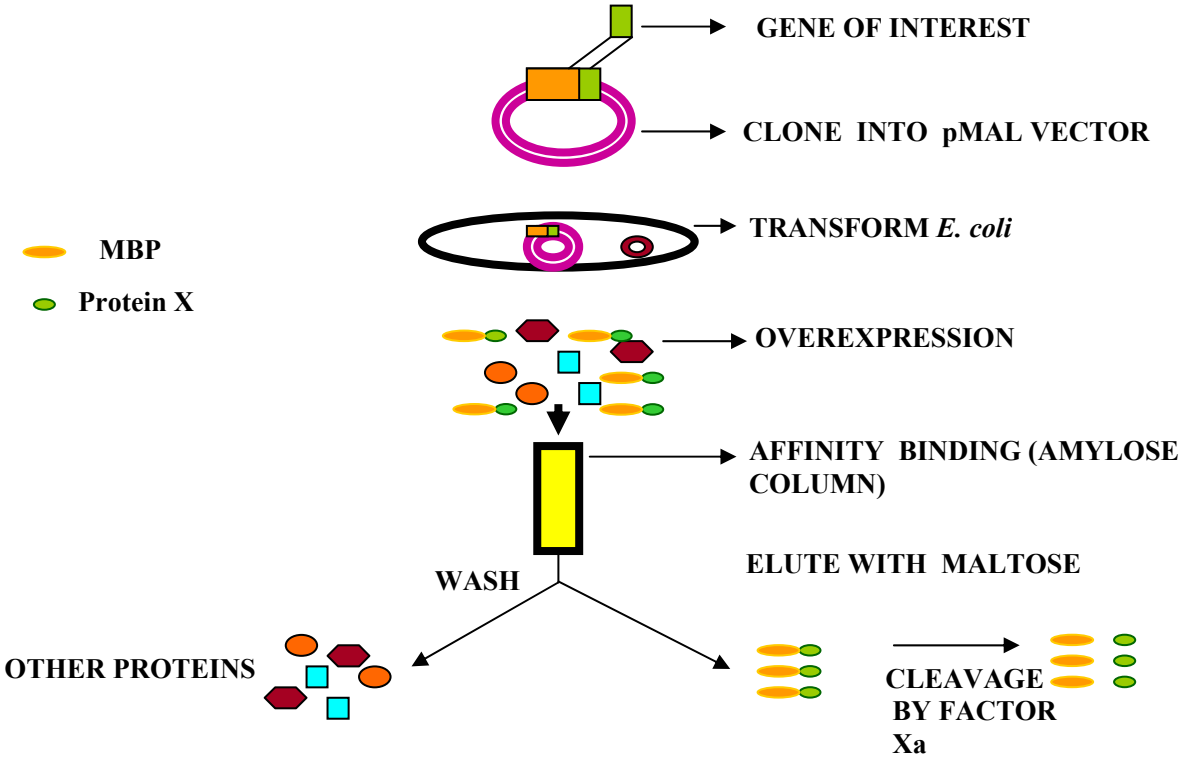



**Figure 2.3 A schematic diagram showing the amylose column purification of the recombinant fusion protein.**

$CaCl_2$ (pH 8.0)] at 23°C for 4-6 hours. Purification and Factor Xa digestion of the recombinant protein was verified by performing 12% SDS-PAGE. SDS-PAGE was performed using prestained low molecular weight markers as protein standards (BioRad, Hercules, California).

**GENERATION OF POLYCLONAL Cah3 AND Cah6 PRIMARY ANTIBODIES**

Factor Xa digested purified recombinant proteins were separated on a 12% gel by SDS-PAGE at 15 mAmp for 18-20 hours. The 29 kDa Cah3 and 31 kDa Cah6 protein bands were excised carefully from the polyacrylamide gel. The gel sections were shipped to Strategic Dissolutions, Ramona, California, for production of antibody. Antibodies were raised against the Cah3 and Cah6 proteins by a standard 70-day protocol using two pathogen free rabbits. Approximately 1.6 mg of the protein was used to raise the antibody.

**CARBONIC ANHYDRASE ASSAYS AND OTHER BIOCHEMICAL EXPERIMENTS**

CA activity was assayed electrometrically using a modification of the Wilbur-Anderson method (Wilbur and Anderson, 1948). The samples were assayed at 4°C by adding 50 to 200 µL of the test sample to 3.5 mL of 20 mM EPPS (4-(2-hydroxyethyl)-1-piperazine propane sulfuric acid; pH 8.0). The reaction was initiated by addition of 1.5 mL of ice cold $CO_2$ saturated water. The time required for the pH drop from 7.7 to 6.3 was measured. The activity of the test sample was calculated using the equation: $WAU = t_o/t\text{-}1$ where t is the time required for the pH change when the test sample is present and $t_o$ is the time required for the pH change when buffer is substituted for the test sample. Bovine CAII (Sigma, St. Louis, Missouri) was used as a positive control. Both Factor Xa cut and uncut purified Cah6, Gclp1 and Cah3 fusion proteins were used for activity assays. Activity assays were also done using protein extracts from induced cells of the γ-CA (Cam) clone of *Methanosarcina thermophila* and purified Cam, which were kindly provided by Dr. J.G. Ferry, Pennsylvania State University, University Park, Pennsylvania. The effect of temperature on the activities of recombinant Cah6 and Cah3

proteins were studied. The temperatures used for this study were 0°C, 4°C, 25°C, 37°C, 45°C and 50°C. Recombinant Cah6 and Cah3 were incubated for 15 minutes at the indicated temperatures and cooled on ice. CA activity was measured at 4°C by the Wilbur-Anderson method. Activity is represented as a percentage of the activity of a sample maintained on ice throughout the experiment.

The effects of the sulfonamide inhibitors ethoxyzolamide and acetazolamide and the anions azide and cyanide on CA activities of recombinant Cah6 and Cah3 were studied. The $I_{50}$ value corresponds to the concentration giving 50% inhibition. $I_{50}$ was determined by plotting the percentage of inhibition vs. the concentration of the inhibitor. Sodium salts of azide and potassium salts of cyanide were used. The effects of the thiol reducing agents like, cysteine, dithiothreitol and 2-mercaptoethanol, on CA activities of recombinant Cah6 and Cah3 were also studied. The recombinant proteins were purified without any thiol reducing agents in the buffer and the specific activities were measured. These specific activities in the oxidized state are denoted as 100% activities (ox-CA). To measure the CA activities in the reduced state, purified recombinant Cah3 and Cah6 enzymes were incubated for 30 minutes at room temperature with 10 mM 2-mercaptoethanol (Mer), 10 mM cysteine (Cys) or 10 mM dithiothreitol (DTT), and then the CA specific activities were measured. All the inhibitors and thiol reducing agents were purchased from Sigma.

**ELECTROPHORESIS AND IMMUNOBLOTTING**

For protein analyses and Western blots, cells were harvested, washed twice with fresh medium and resuspended in TEN buffer (10 mM Tris-HCl, 10 mM EDTA and 150 mM NaCl; pH 7.5). Proteins were separated on 12% and 15% polyacrylamide gels (0.8% bisacrylamide) as described previously (Laemmli, 1970). SDS-PAGE was performed using prestained low molecular weight markers as protein standards (BioRad, Hercules, CA). Immunoblotting was

performed as described in the protocol from BioRad Laboratories (Hercules, California). Cah6 primary antibody was used to probe proteins from high and low $CO_2$ adapted D66 and six different *Cah6* RNAi mutant cells as well as the factor Xa cut and uncut purified MBP-Cah6 fusion proteins. The Cah6 antibody was diluted in the ratio of 1:1000 before being used as a probe.

Both the "new" and the "old" Cah3 primary antibodies were used to probe proteins from air adapted $137^{+}$, *cia3*, *cia3* transformant 9A and *ca-1* cells and also the factor Xa cut and uncut purified MBP-Cah3 fusion proteins. The "old" Cah3 primary antibody was kindly provided by Dr. Göran Samuelsson, Umeå University, Sweden. The "new" and the "old" Cah3 primary antibodies were diluted in the ratio of 1:1000 before being used as a probe. The primary antibodies were diluted with the antibody buffer (Tris-buffered saline + 0.005% Tween 20 + 1% bovine serum albumin, pH 7.4) before being used as probes. The antibody raised against the purified γ-CA from *M. thermophila* was kindly provided by Dr. J.G. Ferry, The Pennsylvania State University, University Park, Pennsylvania. The secondary antibody used for Western blotting was conjugated to the enzyme horseradish peroxidase (BioRad, Hercules, California) and diluted at a ratio of 1:3000 with the antibody buffer. Western blots were developed following the protocol from BioRad using a mixture of the horse radish peroxidase color development reagent (BioRad) in ice cold 100% methanol (20 mL), Tris-buffered saline (80 mL; pH 7.4) and 30% $H_2O_2$ (60 μL).

**IMMUNOLOCALIZATION STUDIES USING ELECTRON MICROSCOPY**

Air adapted D66 cells were fixed in a mixture of 1% $OsO_4$, 2% formaldehyde and 0.5% glutaraldehyde in a 1:1 ratio for 15 minutes. The sample was then fixed for an additional 15 minutes in 1% $OsO_4$, 2% formaldehyde, 0.5% glutaraldehyde and 0.1 mM sodium cacodylate buffer. Materials were rinsed with distilled water and stained with 0.5 % uranyl acetate for 30

minutes. After this, excess stain was rinsed and the samples were dehydrated in ethyl alcohol series. Samples were then infiltrated and embedded in L.R. White resin (Electron Microscopy Sciences, Fort Washington, Pennsylvania). Embedded *Chlamydomonas* tissues were sectioned with a Dupont Sorvall microtome and the sections were 70 μm thick. The immunocytochemical procedure was similar to the method of Borkhsenious *et al*. (1998) with some modifications. Sections were pretreated with 2.5% sodium-meta periodate (Sigma, St. Louis, Missouri) for 10 minutes to remove residual glutaraldehyde, rinsed in distilled water and blocked twice for 20 minutes each with 2% BSA and 0.1% Tween 20 in PBS (Phosphate saline buffer, Sigma).

The sections were then incubated for 90 minutes with diluted primary antibody (1:10 dilutions of the Cah6 primary antibody) or with the preimmune serum diluted similarly (used as a negative control). Sections were then washed with 0.5% Tween 20 six times for a total of 30 minutes and blocked again with 2% BSA(Bovine serum albumin) for 10 minutes. The grids were transferred to 1:50 dilution of Protein A (Sigma) conjugated to 20 nm colloidal gold particles. Protein A was diluted with 1% BSA and 0.1% Tween 20 in PBS for 1 hour. Sections were washed with PBS four times; each rinse was for 5 minutes. Finally the sections were rinsed with distilled water, dried, viewed and photographed by transmission electron microscopy. Sectioning of *Chlamydomonas* cells, immunolocalization and transmission electron microscopy works were done by Ying Xiang (Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana).

**CLONING TO GENERATE CAH6 RNAi MUTANTS**

RNA interference (RNAi) was used in an attempt to generate *Cah6* RNAi mutants. The silencing of a gene triggered by the presence of aberrant RNA of that gene in the host cell is generally termed as “RNA interference”. This abnormal RNA could be a double stranded RNA, a shortened RNA that lacks its “cap” or “tail” or a conventional RNA that is present in unusually

large quantities. All of these may indicate to a cell that a virus is on the attack. The cell responds by degrading these abnormal RNAs. The degradation of these abnormal RNAs is usually accompanied by the suppression of the specific gene in the host cell coding for these abnormal RNAs. Artificially, a RNAi construct of a gene that one wants to silence can be produced by cloning two reverse complementary sequences of the same gene in a vector (Fig. 2.4). This vector can be used to transform a wild type cell. Upon transcription, the RNAi construct leads to the production of double stranded RNA molecules specific to the gene that one wants to underexpress (Fig. 2.4). In the cell, these double stranded RNA molecules are amplified by RNA-dependent RNA polymerase and cleaved into 20-25 bp small fragments. These cleaved double stranded RNA fragments are denatured. The antisense strands from the denatured double stranded RNA molecules bind to the host target mRNA specific to the gene and degrades it. This leads to the lack of expression or underexpression of the target protein which in turn generates the RNAi "mutants" of the target gene. These mutants can be screened under appropriate conditions to study the functional role of the target gene (Fig. 2.4).

Cah6 was cloned into the pSL72 vector which is 5.5 kb long to generate RNAi mutants. The pSL72 vector was kindly provided by Dr. S. D. Lemaire (University of Paris, Orsay, France). The pSL72 vector contains the *AphVIII* and *Amp*$^r$ genes which code for an aminoglycoside phosphotransferase (from *Streptomyces rimosus*) and β-lactamase, respectively. *AphVIII* and *Amp*$^r$ confer paromomycin and ampicillin resistance, respectively. *Amp*$^r$ can be a selective marker in bacteria. *AphVIII* can be a selective marker in *C. reinhardtii* as paromomycin inhibits eukaryotic translation. The *AphVIII* gene has a constitutive PsaD promoter and terminator from *C. reinhardtii*. It also has an intron sequence which is the second intron of the cytochrome $c_6$ gene of *C. reinhardtii* (Fig. 2.5). Two different RNAi constructs (5′ and 3′ end constructs) of *Cah6* were generated by cloning reverse complementary sequences of

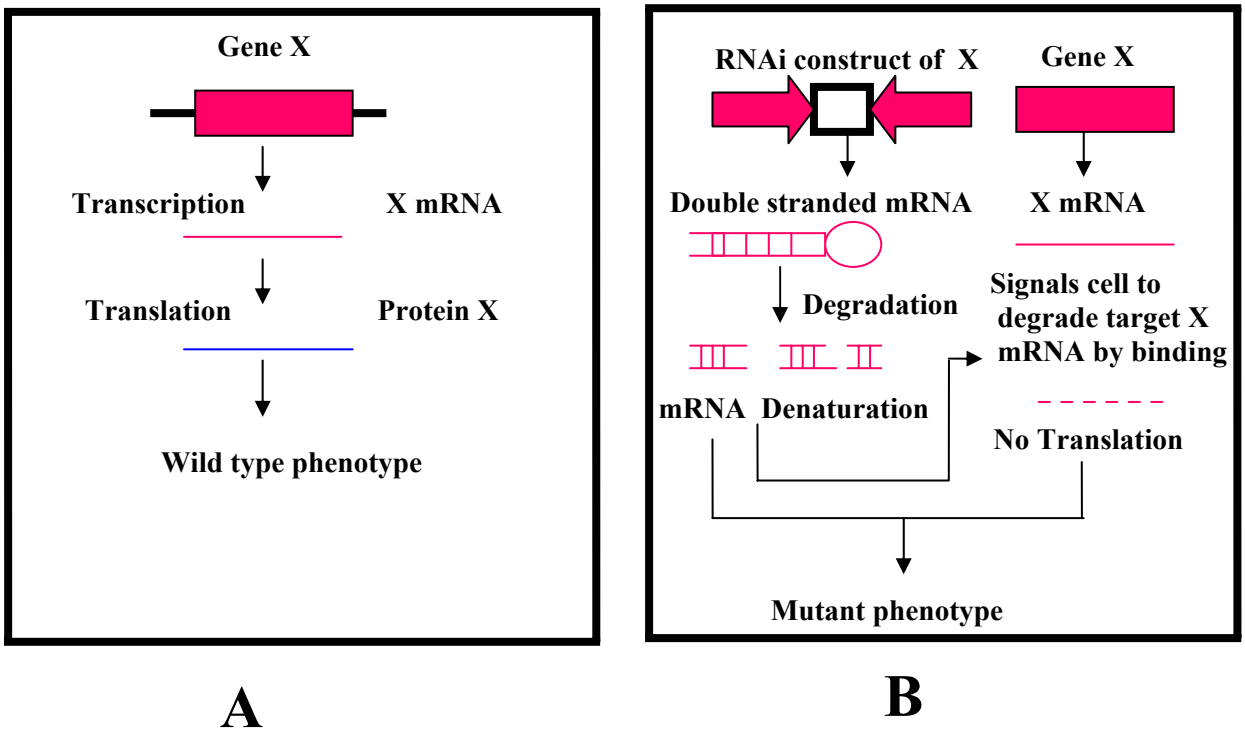



**Figure 2.4 A schematic figure showing the underexpression of the protein by RNA interference.** A. A wild type cell expressing gene X. B. The wild type cell shown in (A) is transformed by the RNAi construct of gene X.

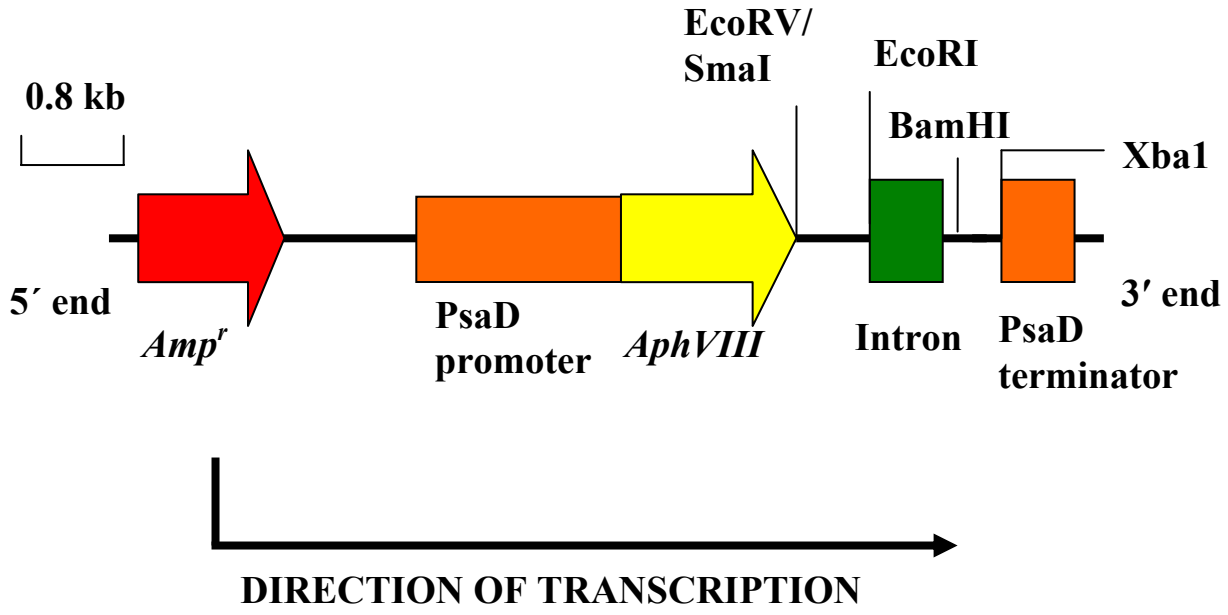



**Figure 2.5 A schematic figure of the pSL72 vector.** The vector is 5495 bp long. *AphVIII* and *Amp*$^r$ genes code for an aminoglycoside phosphotransferase (from *Streptomyces rimosus*) and β-lactamase respectively. *AphVIII* and *Amp*$^r$ confer paromomycin and ampicillin resistance, respectively. The PsaD promoter and terminator are from *C. reinhardtii*. The intron sequence is the second intron of the cytochrome $c_6$ gene of *C. reinhardtii*. The restriction enzyme sites that were used for cloning are given on the top of the construct by black lines.

Cah6 in the pSL72 vector (Fig 2.6). The 5′ end of the construct contained both genomic and cDNA sequences of a selected 5′ end region of the *Cah6* gene that were reverse complementary to each other. RNAi primers X1 + X2 (Appendix 2) were designed to amplify a product of 702 bp of *Cah6* from the cosmid 72-E-6. The genomic product had two introns of *Cah6*. X3 + X4 primers (Appendix 2) were used to amplify a product of 457 bp of *Cah6* from the cDNA core library which is reverse complementary to X1 + X2 PCR product. The 3′ end construct had selected reverse complementary 3′ UTR cDNA sequences of *Cah6* (Fig 2.6). Sequences of these RNAi primers and their alignments on the genomic or cDNA map are shown in Appendix 2. The primers X5 + X6 and X7 + X8 (Appendix 2) were designed to amplify the two 609 bp reverse complementary cDNA PCR products that were used to generate the 3′ end construct. The 5′ and 3′ end regions of *Cah6* were selected based on comparative DNA sequence analyses of the mitochondrial β-CAs (*Ca1* and *Ca2*) and *Cah6* genes.

Reverse complementary sequences were cloned in two steps. In the first round of cloning, PCR products generated by X1 + X2 and X5 + X6 primers were digested by EcoRI and SmaI. The vector was double cut with the EcoRI and EcoRV. PCR products (X1 + X2 and X5 + X6) were ligated separately before the 5′ end of the cytochrome $c_6$ intron in the pSL72 vector to generate 5′ (clone # 6) and 3′ (clone # 3) end preliminary RNAi constructs, respectively (Fig. 2.6). *Cah6* transformants containing part of the *Cah6* gene were screened on LB + Amp (100 μg $mL^{-1}$) plates and positive clones were verified by restriction digestion with AvrII and EcoRI. In the second cloning step, PCR products generated by X3 + X4 and X7 + X8 primers were digested with Bam HI (at the 5′ end) and XbaI (at the 3′ end) and separately ligated to the 5′ (clone # 6) and 3′ (clone # 3) end preliminary RNAi constructs, respectively obtained in the first round of cloning. The *Cah6* DNA fragments (X3 + X4 and X7 + X8) in the second round of cloning, were cloned after the 3′ end of the cytochrome $c_6$ intron in pSL72 vector (Fig. 2.6).

Transformants were screened on LB + Amp plates as before. Double restriction enzyme digestions of the selected clones with various combinations of restriction endonucleases were done (EcoRI and AvrII, BamHI and XbaI, Bam HI and AvrII and AvrII and XbaI). These restriction digestion results were verified by DNA sequencing. This led to the identification of the 5′ (clone # 11A) RNAi and 3′ (clone # 10B) RNAi clones. Bacterial transformation was done according to the protocol in Sambrook *et al.*, (1989) and the ligations were done using T4 DNA ligase from NEB, following their protocol.

**TRANSFORMATION OF *C. reinhardtii* AND SCREENING OF RNAi MUTANTS**

The strain D66 (nit2$^-$, cw15, mt$^+$) was used for transformation. Culture conditions were similar to those used previously (Rawat and Moroney, 1991). D66 cells were first grown on 100 mL of Tris-Acetate-Phosphate (TAP) medium (Sueoka, 1960) for 2 days and then transferred to 1 L of TAP medium 24 hours before the experiment. Cells were harvested and resuspended at a density of 2 x $10^8$ cells mL$^{-1}$. For electroporation, 1 μg of circular or linearized DNA of 5′ and 3′ Cah6 RNAi constructs were added to 300 μL of the resuspended cells in electroporation cuvettes (0.4 cm gap width; BioRad Laboratories, Hercules, California). DNA was linearized by digestion with KpnI. These mixtures were placed on ice for 15 minutes. A BioRad electroporator Gene pulser II at a capacitance of 25 μF, without a shunt resistor, a voltage set at 2000 V cm$^{-1}$ and a pulse time between 9.4 and 10 ms, was used for electroporation. After electroporation, the cells were allowed to recover overnight in 10 mL of TAP + 60 mM sucrose medium in the dark. The next morning, the cells were harvested and resuspended in 4 mL of TAP + 0.5% agar. Cells were plated onto TAP + paromomycin (7.5 μg mL$^{-1}$) plates and were allowed to grow under very low-light conditions. Transformants containing the RNAi constructs had the paromomycin resistance gene and grew on TAP + paromomycin plates. Three hundred 5′ RNAi and 600 hundred 3′ RNAi paromomycin resistant colonies were transferred to

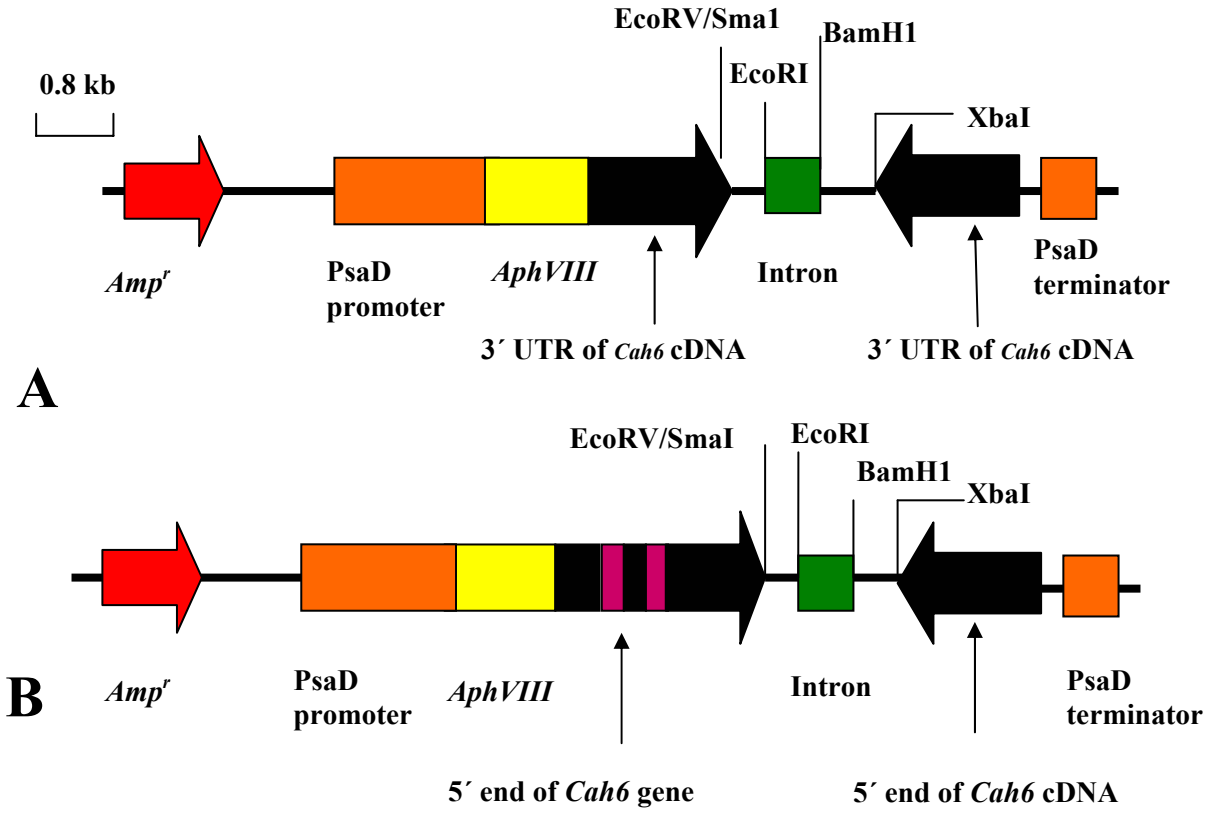



**Figure 2.6 RNAi constructs of *Cah6*.** A. 3′end *Cah6* RNAi construct. B. 5′end *Cah6* RNAi construct The reverse complementary fragments of *Cah6* were cloned in the pSL72 vector. Restriction enzymes used to cut the vector and the Cah6 DNA fragments are shown by straight black lines. *AphVIII* and *Amp*$^r$ code for paromomycin and ampicillin resistance, respectively. PsaD promoter and terminator are from *C. reinhardtii*. PsaD promoter is a constitutive promoter. The intron represented by the green box is the second intron of the cytochrome $c_6$ gene of *C. reinhardtii*. Two pink boxes in the 5′ end *Cah6* RNAi construct represent the two introns in the 5′ end of *Cah6* gene.

fresh TAP-paromomycin plates. After one week, 5-6 colonies from the 5′ RNAi and 3′ RNAi plates were randomly picked and screened under high and low $CO_2$ conditions on MIN plates for the presence of Cah6 protein using Western blotting.

**OTHER ANALYTICAL METHODS**

The $CO_2$ concentration in the growth chambers was measured using an infrared gas analyzer (The Analytical Development Co. Ltd, Hoddesdon, England) which reads at an accuracy of ± 2%. The $CO_2$ concentration was checked at least once, everyday, while the alga was growing in the high and low $CO_2$ growth chambers. Protein concentration was determined by the method of Lowry *et al*. (1951) with bovine serum albumin as standard. Chlorophyll concentration was determined spectrophotometrically (Arnon, 1949) using the equations of Holden (1976). 100% methanol was used as a solvent for extraction of chlorophyll. Cell density values were determined by direct counting in a hemocytometer chamber.

# CHAPTER 3

# OVEREXPRESSION, PURIFICATION AND PARTIAL CHARACTERIZATION OF THE α-CA PROTEIN Cah3

## INTRODUCTION

Of the three CA families, the α-CA is the most studied and has a very wide distribution. Although α-CAs have long been known to occur in animals (Meldrum and Roughton, 1933), they have only recently been identified in algae (Fukuzawa *et al*., 1990; Fujiwara *et al*., 1990), the higher plant *Arabidopsis thaliana* (Moroney *et al*., 2001) as well as in eubacteria (Soltes-Rak *et al*., 1997; Elleby *et al*., 2001; Chirica *et al*. 2001) and viruses (Niles *et al*., 1986; Strayer and Jerng, 1992). Most α-CAs are active as monomers of about 30 kDa with three histidines coordinating the zinc atom (Moroney *et al*. 2001). All known α-CAs are highly susceptible to inhibition by sulfonamide compounds (Moroney *et al*., 1985). More information about α-CAs may be found in Chapter 1.

Three α-CAs have been identified to date in the green alga *C. reinhardtii*. In *C. reinhardtii*, two of these α-CAs (Cah1 and Cah2) are compartmentalized in the periplasm and one (Cah3) occurs in the lumen of the thylakoid. *Cah1* and *Cah2* were the first α-CA genes cloned from a photosynthetic organism (Fukuzawa *et al*., 1990; Fujiwara *et al*., 1990). These two genes encode very similar proteins although they are regulated differently. *Cah1* is expressed under low $CO_2$ conditions but not under high $CO_2$ conditions while *Cah2* is poorly expressed under low $CO_2$ and slightly upregulated under high $CO_2$ conditions. In addition, the expression of *Cah2* under high $CO_2$ appears low compared to the expression of Cah1 under low $CO_2$ conditions (Fujiwara *et al*., 1990). A HCR mutant, *cia5,* fails to produce the low $CO_2$-induced CA and any of the proteins associated with the CCM. These results seem to indicate

that Cah1 is required for the functional operation of CCM in *C. reinhardtii* (Moroney *et al*., 1985). Recent studies have cast some doubt on the on the role of Cah1 since a *Chlamydomonas* mutant defective in the *Cah1* gene and protein appears to perform as well as wild type cells at limiting $C_i$ conditions (Van and Spalding, 1999). Thus at this point the evidence for the role of Cah1 in the CCM is not very clear. The function of Cah2 is also obscure (Moroney *et al*., 2001).

Cah3 from *C. reinhardtii* was discovered in 1995 (Karlsson *et al*., 1995). This α-CA is constitutively expressed and is slightly upregulated under low $CO_2$ conditions. A *C. reinhardtii* mutant, *cia3*, which has two point mutations in the region coding for the transit peptide of the Cah3 protein, grows poorly under low $CO_2$ but grows well under high $CO_2$. *cia3* and *ca-1* (another HCR mutant), were complemented by the wild type *Cah3* gene (Karlsson *et al*., 1998; Funke *et al*., 1997) and the complemented cells grew well under low $CO_2$. This shows that Cah3 is required for the growth of *C. reinhardtii* in air levels of $CO_2$ (Funke *et al*., 1997). This also shows that the mutation in both mutants affects the same gene. The sequencing of the genomic region of *ca-1* revealed a point mutation in the 5′ end of the ORF (open reading frame) creating a stop codon. Translation of the Cah3 mRNA in *ca-1* led to an incomplete and inactive polypeptide which was thirteen amino acids long. Karlsson *et al*. (1998) published the sequence of the *Cah3* cDNA. This group had cloned the Cah3 cDNA sequence coding for the mature Cah3 protein in the pMal-c2x overexpression vector. They had fused the Cah3 cDNA at the 3′ end of the *MalE* gene in this vector. MalE codes for the maltose binding protein (MBP). More information about the cloning and the MalE-*Cah3* recombinant construct has been presented in Chapter 2.

The *MalE-Cah3* recombinant construct and the Cah3 primary antibody raised against the purified uncut MBP-Cah3 protein were kindly provided for my research by Dr. Göran Samuelsson (Umeå University, Sweden). Here I report my studies of the Cah3 protein.

Although the *Cah3* gene has been well characterized, there is no past report of the detection and measurement of CA activity of the purified protein. This is the first report of detection of significant *in vitro* CA activity in a purified Cah3 protein and partial characterization of this CA activity.

**RESULTS**

**1. Overexpression of the MBP-Cah3**

The original Cah3 antibody ("old") was generated by Dr. Samuelsson's laboratory (Umeå University, Sweden) against the purified MBP-Cah3 recombinant fusion protein (Karlsson *et al*., 1998). This "old" antibody did not cross react with Cah3 specifically. Hence Cah3 was overexpressed and purified primarily to raise an antibody against it for Western blotting and immunolocalization studies. *E. coli* cells harboring the recombinant *MalE-Cah3* construct (Fig. 3.1) were induced with 1 mM IPTG for 2 hours at 37°C to overexpress the MBP-Cah3 fusion protein (Chapter 2). The overexpressed recombinant Cah3 protein comprised approximately 10% of the total *E. coli* cell protein (Fig. 3.2).

**2. Purification and Activity Assays of the Recombinant Cah3**

Overexpressed recombinant Cah3 protein was purified by affinity chromatography using amylose resin. The purified fusion protein was further concentrated by using 100 kDa cut-off centricon columns. At each step of purification of the recombinant Cah3 protein from crude *E. coli* cell sample, the CA activity in the sample was assayed in order to check the purity of the Cah3 sample (Table 3.1). The purified recombinant Cah3 had a specific activity of 300 WAU/mg (Wilbur and Anderson unit/mg). This calculation of the specific activity of Cah3 was based on the total amount of recombinant fusion protein in the sample. CA activity was not detected in either the cell extracts from uninduced *E. coli* cells harboring the recombinant *MalE-Cah3* plasmid or *E. coli* cells containing only the pMal vector. The fusion protein was cleaved by

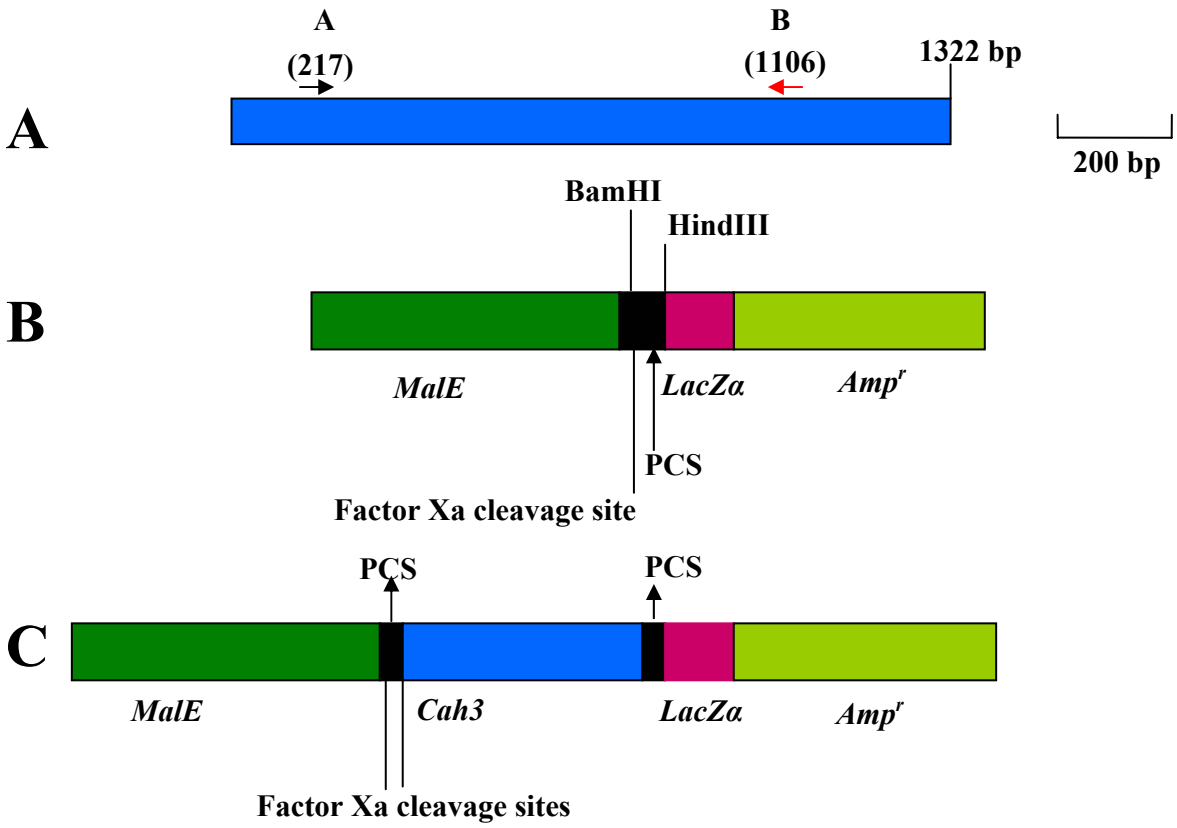



**Figure 3.1 A schematic diagram of the recombinant *MalE-Cah3* expression construct.** A. schematic figure showing the alignment of primers used to amplify the *Cah3* cDNA. A and B are 5′ and 3′ end PCR primers, respectively for amplification of the *Cah3* cDNA. Primer B has a HindIII site incorporated at the 5′ end. Primer A has a BamHI and a Factor Xa site incorporated at the 5′ end. The numbers on the top denote the primer positions in base pairs. B. Part of the pMal vector showing the polylinker cloning site (PCS), β-galactosidase (*LacZα*), β-lactamase (*Amp*$^r$) genes, BamHI and HindIII recognition sites and Factor Xa cleavage sites. *MalE* codes for the MBP. C. A schematic figure of the *Cah3* recombinant construct.

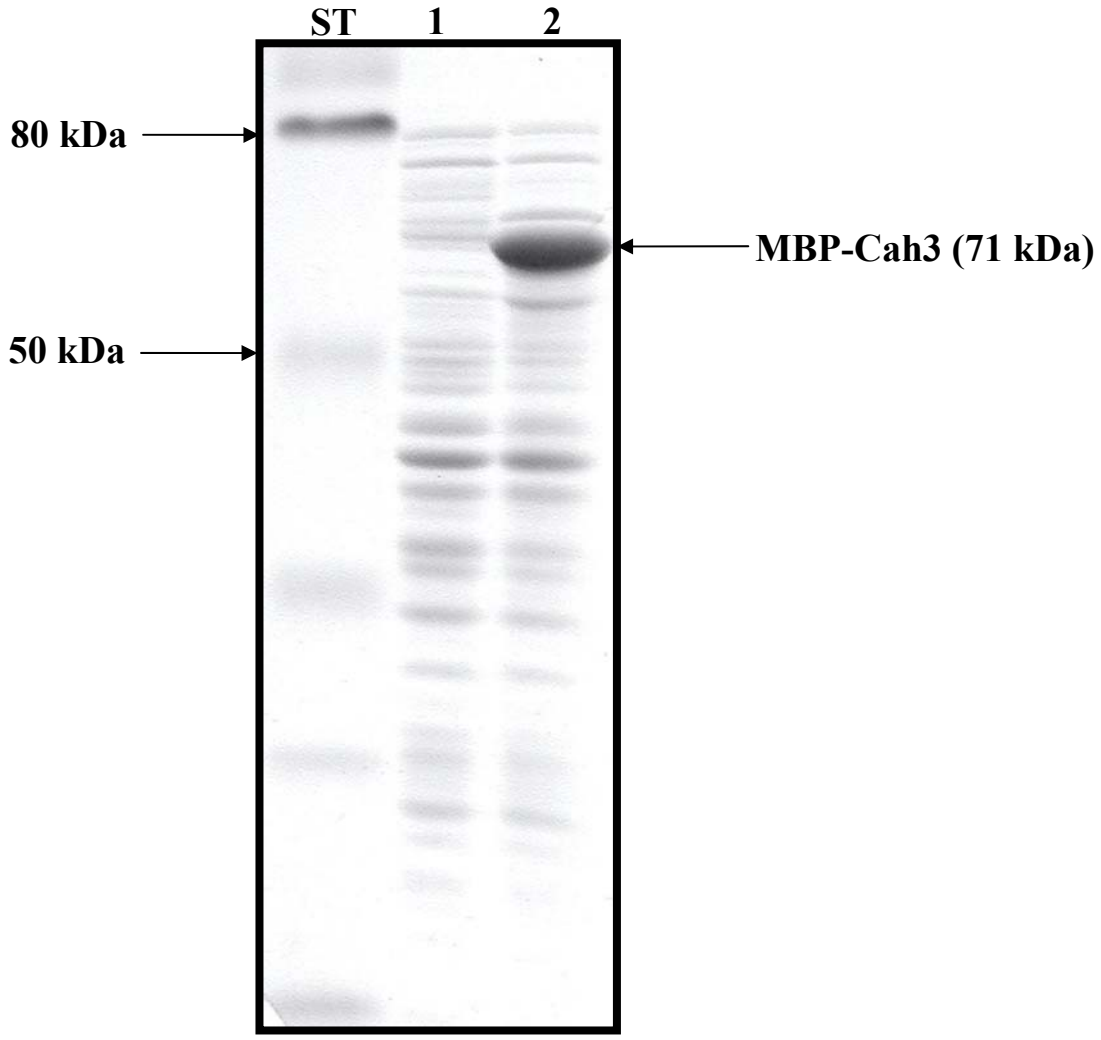



**Figure 3.2 A 12% SDS-polyacrylamide gel showing the uninduced and induced *Escherichia coli* cells expressing the recombinant MBP-Cah3 protein.** Lane ST contains prestained molecular weight markers. Lane 1 contains 20 µg of proteins from uninduced *E. coli* cells. Lane 2 contains 20 µg of proteins from induced *E. coli* cells expressing chimeric MBP-Cah3 protein. Proteins were stained with Coomassie Blue after 12% electrophoresis.

**Table 3.1** The purification of the chimeric MBP-Cah3 from 2 liters of *E. coli* culture

| Step | Total activity (WAU)[a] | Protein (mg)[b] | Specific activity (WAU/mg) | Recovery (%) | Purification fold |
|---|---|---|---|---|---|
| Sonicated cells | 1000 | 8333 | 0.12 | 100 | 1 |
| Amylose column | 800 | 5 | 160 | 80 | 1334 |
| Centricon column | 300 | 1 | 300 | 30 | 2500 |

[a] one WAU = $(t_0-t)/ t$ where $t_0$ is the time for uncatalyzed reaction and t is the time for the enzyme catalyzed reaction; [b] determined by Lowry's protein assay (Lowry *et al*., 1951)

the protease Factor Xa for four hours at 23°C to cleave Cah3 from the MBP. Purification and cleavage of the fusion protein was confirmed by SDS-PAGE (Fig. 3.3). The 29 kDa Cah3 protein band was excised from the gel to be used as a source antigen for production of the new polyclonal Cah3 primary antibody ("new").

**3. Western Blotting**

The original Cah3 antibody ("old") was generated by Dr. Samuelsson's laboratory (Umeå University, Sweden) against the purified MBP-Cah3 recombinant fusion protein (Karlsson *et al.*, 1998). Factor Xa-cleaved Cah3 and purified MBP-Cah3 fusion proteins were separated by 12% SDS-PAGE and probed with the "old" and the "new" Cah3 primary antibodies (Fig. 3.4 and Fig. 3.5). The "old" Cah3 antibody reacted equally well with the MBP and Cah3 (Fig. 3.4). The new Cah3 antibody reacted specifically with the Cah3 protein since it was separated from the MBP from the gel before raising the antibody (Fig. 3.5).

Proteins were extracted from cells (grown at air levels of $CO_2$) of different strains of *C. reinhardtii*, namely WT 137, *ca-1* (*cah3* mutant), *cia3* (*cah3* mutant) and 9A (a *cia3* mutant transformed with the wild type *Cah3* gene). These proteins were separated by 10% SDS-PAGE and probed with both the "new" and the "old" Cah3 primary antibody (Fig. 3.6). The "new" Cah3 antibody detected the Cah3 protein (29 kDa band) in the WT 137 and 9A cells but not in the two *cah3* mutants namely *ca-1* and *cia3*. Moreover, the 9A transformant cells possessed approximately twice the amount of Cah3 protein than that of the WT 137 (Fig. 3.6). The "old" Cah3 antibody picked up two distinct proteins having very similar molecular weights. The top and the bottom protein bands picked up by the "old" antibody are approximately 30 kDa and 29 kDa, respectively. The bottom protein band (29 kDa) matches the 29 kDa protein band detected by the "new" Cah3 antibody and is that of Cah3 protein. This 29 kDa Cah3 protein was missing in the *ca-1* and *cia3* mutants but present in WT137 and 9A transformant. This 30 kDa band was

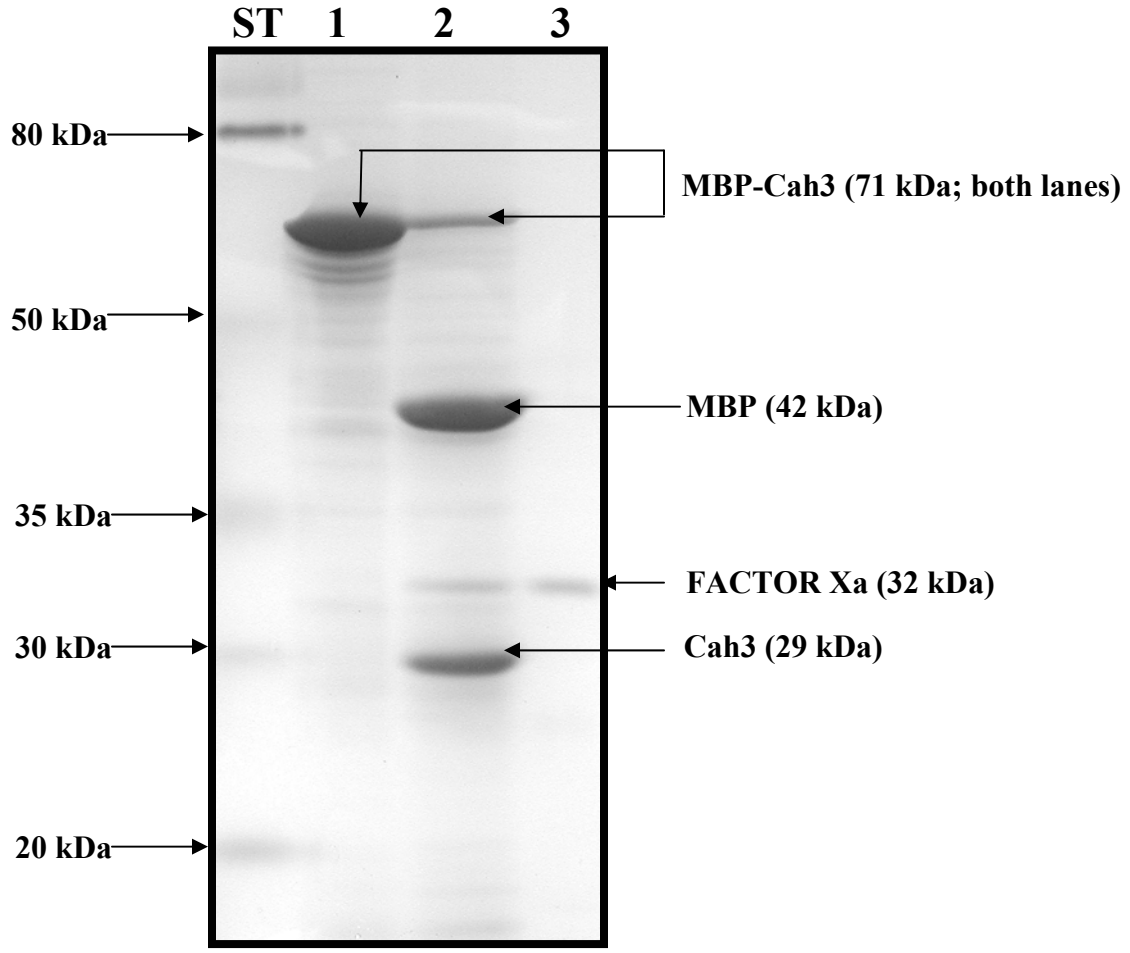



**Figure 3.3 A 12% SDS-polyacrylamide gel showing the purified uncut and the Factor Xa cut MBP-Cah3 protein.** Lane ST represents the prestained molecular weight markers. Lane 1 contains 20 μg of uncut chimeric MBP-Cah3 protein. Lane 2 contains 20 μg of chimeric MBP-Cah3 cut by 1 μg of Factor Xa. Lane 3 contains 1 μg of Factor Xa. Proteins were stained with Coomassie Blue after 12% electrophoresis.

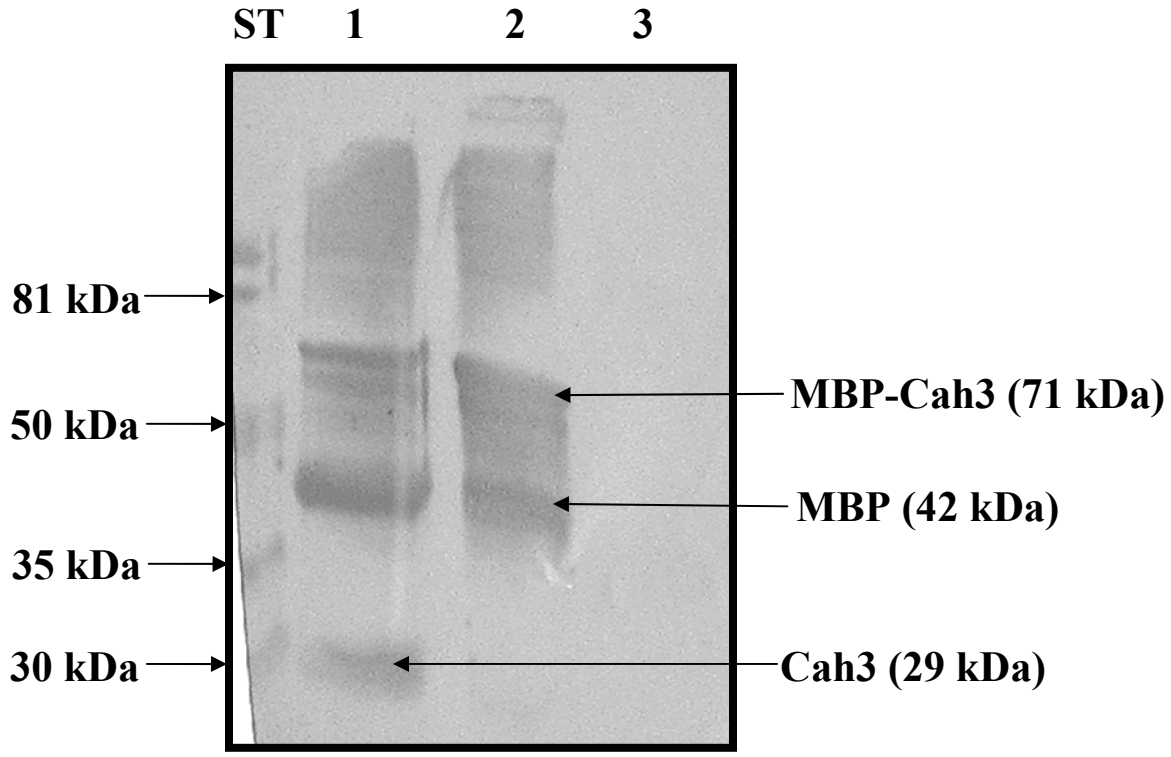



**Figure 3.4 A Western blot probed by the "old" Cah3 antibody raised against the recombinant mature MBP-Cah3 protein.** ST represents the low molecular weight prestained standards. Lane 1 contains 25 μg of purified MBP-Cah3 cut with 1 μg of Factor Xa. Lane 2 contains 25 μg of purified uncut MBP-Cah3. Lane 3 contains 1 μg of Factor Xa. The proteins were run on a 12% SDS-polyacrylamide gel followed by Western blotting.

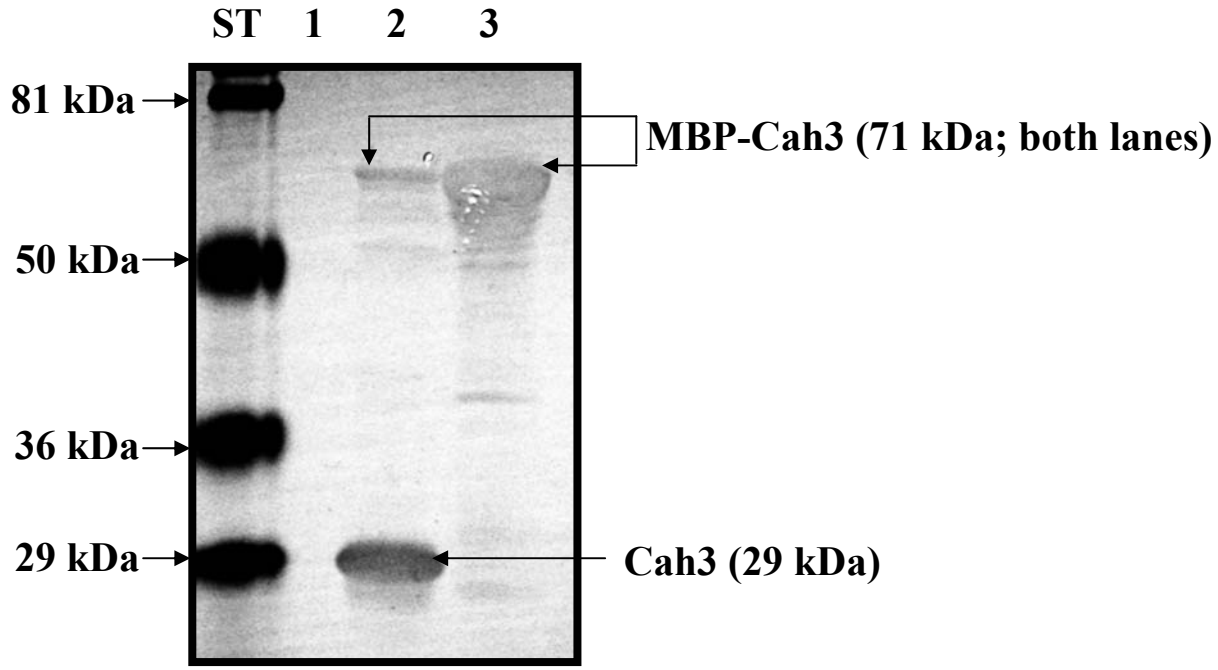



**Figure 3.5 A Western blot probed by the "new" Cah3 primary antibody.** Lane ST represents the prestained molecular weight markers. Lane 1 contains 1 μg of Factor Xa. Lane 2 contains 20 μg of purified chimeric MBP-Cah3 cut by 1 μg of Factor Xa. Lane 3 contains 20 μg of purified uncut chimeric MBP-Cah3.

picked up by the "old" antibody when it was used to probe proteins extracted from spinach leaves (data not shown). The identity of the top 30 kDa band is unknown at this point. These two proteins were not resolved on a 12% or 15% gel as they migrated together on the gel (Fig. 3.7). A number of labs (Park *et al*., 1999; Villarejo *et al*., 2001; Villarejo *et al*. 2002; Karlsson *et al*., 1998) have used this "old" Cah3 antibody for research, and the mutant *cia3* has been reported to have the Cah3 protein, when the "old" Cah3 antibody was used as a detecting tool (Karlsson *et al*., 1998). There is currently no published report stating that *ca-1* has the protein.

**4. Partial Characterization of the Cah3 Activity**

Since there appears to be no significant difference between the catalytic activity of Factor Xa-cleaved MBP-Cah3 and the fusion protein, all the biochemical experiments were done with the purified MBP-Cah3 fusion protein. Since sulfonamide compounds and certain anions are known to inhibit CA activity (Johansson and Forsman, 1993), the effects of sulfonamide and anion inhibitors on the CA activity of recombinant Cah3 were studied (Table 3.2). Sulfonamides were more potent as inhibitors than monovalent anions (cyanide and azide). Ethoxyzolamide was a more effective inhibitor than acetazolamide. Azide was slightly more effective as an inhibitor than cyanide.

The effects of different temperatures on the Cah3 CA activity are shown in Fig. 3.8. The enzyme lost more than 50% of its activity above 43°C and became inactive at 50°C. The sensitivity of recombinant Cah3 to reduction was studied. Different SH-reducing agents were used to reduce Cah3 and observe the reduction effects on the CA activity of recombinant Cah3. Cah3 lost most of its CA activity upon exposure to SH-reducing agents. Purification of the enzyme with buffers lacking a SH-reducing agent yields an enzyme with high activity as shown in Fig. 3.9. The result seems to indicate that the enzyme is more active in the oxidized state than the reduced state.

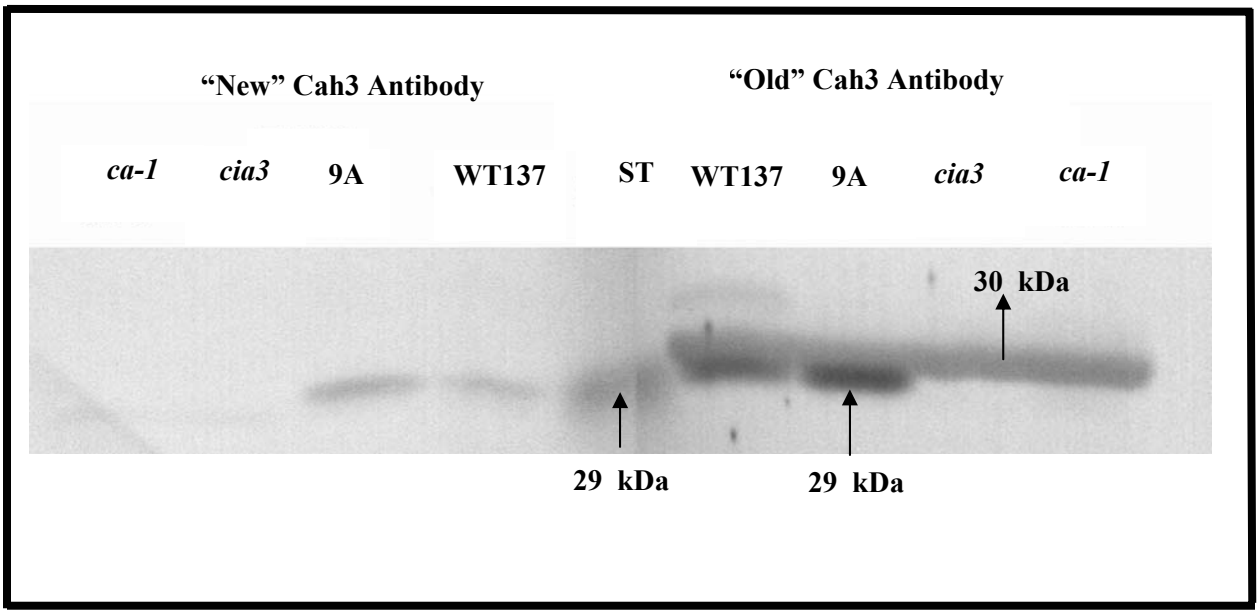



**Figure 3.6 Western blots probed by the "old" and "new" Cah3 primary antibodies when samples were separated by 10% SDS-PAGE.** *ca-1* and *cia3* are two different *cah3* mutants of *C. reinhardtii*. 9A is *cia3* mutant transformed with the wild type *Cah3* gene. WT 137 represents the parental wild type cell of *cia3*. ST represents the prestained molecular weight markers.

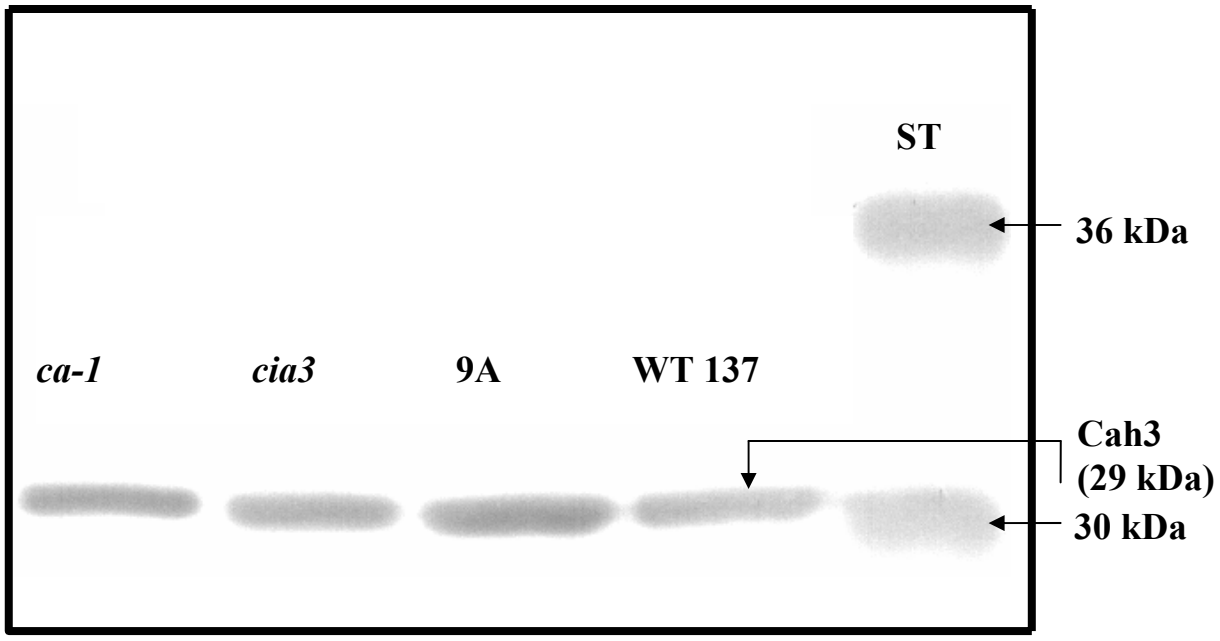



**Figure 3.7 Western blots probed by the "old" Cah3 primary antibody when samples were separated by 15% SDS-PAGE.** *ca-1* and *cia3* are two different *cah3* mutants of *C. reinhardtii*. 9A is *cia3* mutant transformed with the wild type *Cah3* gene. WT 137 represents the parental wild type cell of *cia3*. ST represents the prestained molecular weight markers.

**Table 3.2** The inhibition constants of the bovine CAII and heterologously produced Cah3

| Inhibitor | $I_{50}$ of bovine CAII (M) | $I_{50}$ of Cah3 (M) |
|---|---|---|
| Acetazolamide | $1.4 \times 10^{-8}$ | $8 \times 10^{-9}$ |
| Ethoxyzolamide | $1.2 \times 10^{-9}$ | $6 \times 10^{-9}$ |
| Azide | $1.1 \times 10^{-3}$ | $32 \times 10^{-6}$ |
| Cyanide | $4.9 \times 10^{-5}$ | $59 \times 10^{-6}$ |

Erythrocyte Bovine CAII was purchased from Sigma. The $I_{50}$ value corresponds to the concentration giving 50% inhibition. $I_{50}$ was determined by plotting percentage of inhibition vs. concentration of inhibitor. Sodium salts of azide and potassium salts of cyanide were used. The data shown for each treatment in the table are the averages ± SD of three different measurements.

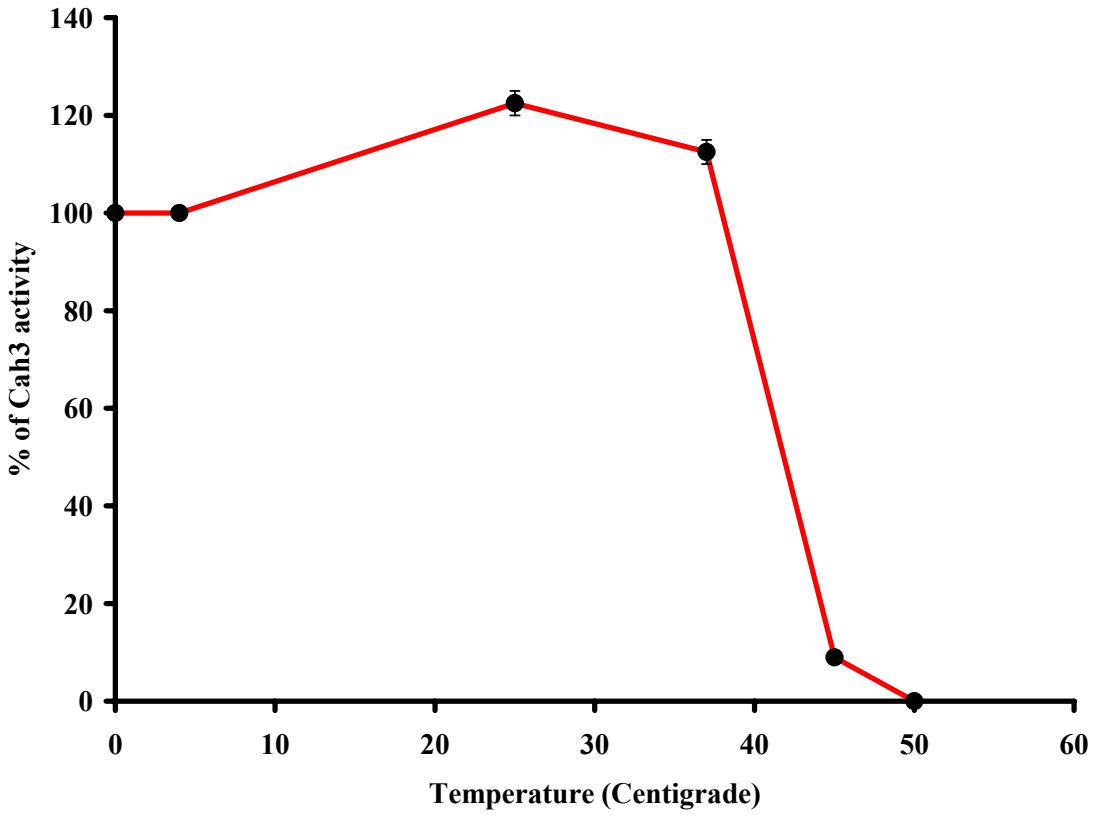



**Figure 3.8 The thermostability of Cah3 activity.** Cah3 was incubated for 15 minutes at the indicated temperatures and cooled on ice. CA activity was measured at 4°C by the Wilbur-Anderson method (1948). Activity is represented as a percentage of the activity of a sample maintained on ice throughout the experiment (300 WAU/mg). The data shown for each temperature treatment in the graph are the averages ± SD of three different measurements.

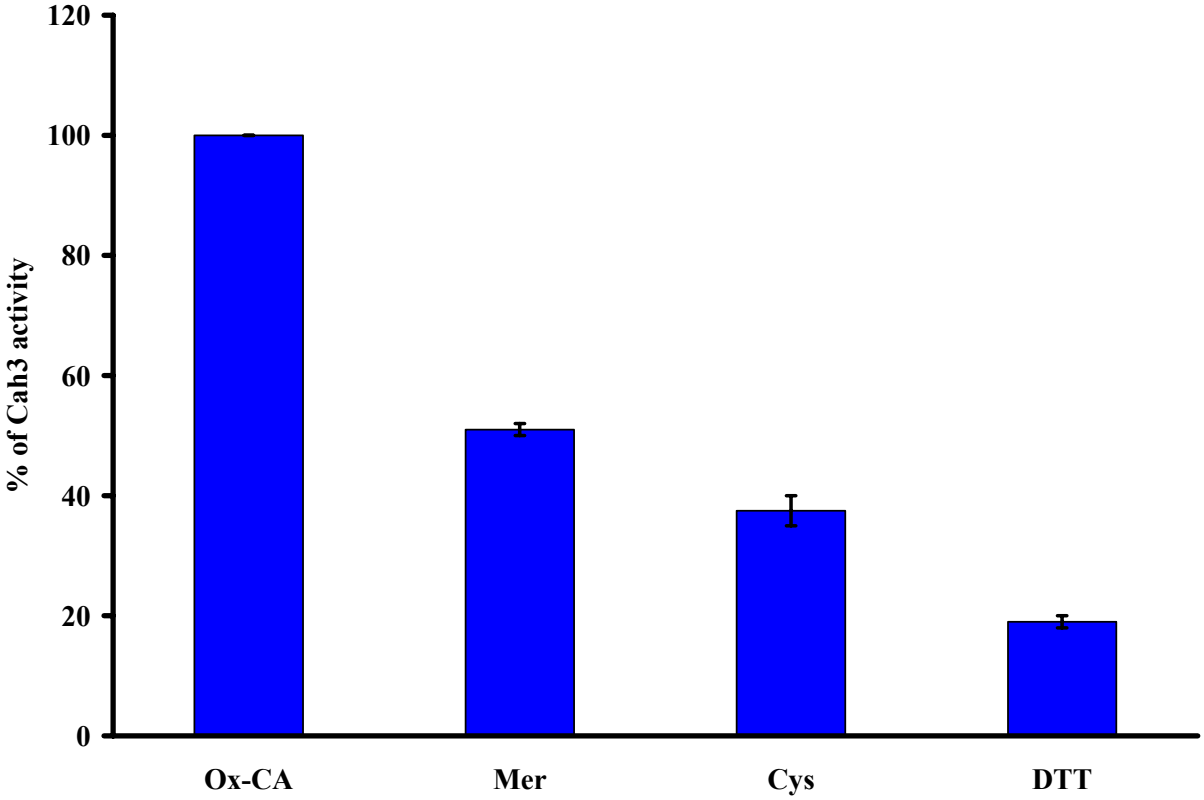



**Figure 3.9 The effects of the SH-reducing agents on Cah3 activity.** Cah3 was purified without any reducing agents in the buffer and the activity was measured. The specific activity (300 WAU/mg) in the oxidized state is denoted as 100% activity (ox-CA). Prior to the measurements of CA activity, purified recombinant Cah3 enzyme was incubated for 30 minutes at room temperature with: 10 mm 2-mercaptoethanol (Mer), 10 mM Cysteine (Cys) and 10 mM dithiothreitol (DTT). The data shown for each treatment in the graph are the averages ± SD of three different measurements.

## DISCUSSION

This is the first report of the detection of CA activity in purified recombinant Cah3 and partial characterization of this CA activity. Cah3 is an α-CA located in the thylakoid lumen. This α-CA is constitutively expressed and slightly upregulated under low $CO_2$ (Karlsson *et al*.,1998). Cah3 has 30%- 40% amino acid identity with other known α-CAs and is most similar to the α-CA from *Neisseria gonorrhoeae* [(40.6% identity); Karlsson *et al*., 1998]. α-CAs are known to have a broad range of specific activity ranging from 5 WAU/mg (human CAIII) to 2000 WAU/mg (human CAII). The specific activity of recombinant Cah3 enzyme is around 300 WAU/mg, comparable to human CAI (an α-CA) specific activity. The calculation of the specific activity in the present study was based on the total amount of recombinant fusion protein in the sample, and MBP and Cah3 have a molecular weight of 42 kDa and 29 kDa, respectively. Since the molecular size of MBP is 1.44 times greater than that of Cah3, the actual specific activity of Cah3 would be approximately 730 WAU/mg. The recombinant mature Cah3 was heterologously produced in *E. coli*. It is known that proteins expressed in bacterial cells sometimes are not folded properly or mostly get aggregated in inclusion bodies. Significantly, less than 10% of the total purified Cah3 protein was in the soluble fraction and was active. This is evident from the extent of purification and specific activities (shown in Table 3.1), and the SDS-PAGE results (Fig. 3.2), respectively. Cah3 is targeted to the thylakoid lumen which has an acidic pH around 4. Cah3 might have an optimum pH around 4. CA activity assays were done at pH 8 as the Wilbur and Anderson CA activity assay method works best around pH 8.5-pH 7.4. This could have had an effect on the Cah3 CA activity.

Cah3 has an apparent molecular weight of 29 kDa on a SDS-polyacrylamide gel and a pI of ~ 7.87. The calculated molecular weight of mature Cah3 is around 26 kDa. It is not a membrane spanning protein as it can be released into the soluble fraction by washing the

thylakoids with 300 mM KCl (Karlsson *et al*., 1998). It was initially suggested that Cah3 speeds up the formation of $CO_2$ from $HCO_3^-$ in the acidic lumen of thylakoids and that this $CO_2$ diffuses through the thylakoid membrane to the pyrenoid (Moroney and Mason, 1991; Badger and Price, 1994; Raven, 1997; Moroney and Somanchi, 1999). This model is based on the assumption that $HCO_3^-$ is actively pumped into the lumen from the stroma and on the observation that in many algae, including *Chlamydomonas*, thylakoids penetrate the pyrenoid matrix in various patterns (Raven, 1997).

It has also been proposed that the activity of this CA stabilizes the manganese cluster of PSII and maintains the oxygen evolution complex (OEC) in a functionally active state (Villarejo *et al*., 2002). Recently, Hanson *et al*., (2003) have proposed that Cah3 mainly provides elevated levels of $CO_2$ to Rubisco under low $CO_2$ conditions and under high light, and might indirectly aid in the stabilization of PSII by preventing overacidification of the thylakoid lumen through rapid dehydration of $HCO_3^-$ to $CO_2$ and $H_2O$.

Cah3 has three histidine residues that coordinate the zinc atom, and is highly susceptible to sulfonamide compounds like all other α-CAs. Recombinant Cah3 is more susceptible to acetazolamide, azide and cyanide and slightly less susceptible to ethoxyzolamide than bovine CAII. Membrane impermeant inhibitors of CAs like acetazolamide (AZ) inhibits $CO_2$ fixation and $C_i$ accumulation in wild type *C. reinhardtii* cells (Moroney *et al*., 1985). Ethoxyzolamide (EZ) is a membrane permeable CA inhibitor. When physiological studies were performed using EZ on the air adapted wild type *Chlamydomonas* cells, the cells showed the same physiological properties as the air adapted *cia3* cells (Spalding *et al*., 1983; Karlsson *et al*., 1998). These physiological traits include reduction in internal CA activity, a high photosynthetic $CO_2$ compensation point, oxygen sensitive photosynthesis, high rates of glycolate accumulation and excessive accumulation of $HCO_3^-$ compared to the wild type cells. These observations suggest

that both the periplasmic and thylakoid α-CAs (Cah1 and Cah3) play an important role in CCM and photosynthesis.

Cah3 loses more than 50% of its activity at temperatures at 43°C and 100% of its CA activity at 50°C. It has an optimal activity around 33°C. Thus the optimal temperature for Cah3 activity falls within the range of the optimal growth temperature (12°C-35°C) of *C. reinhardtii*. The enzyme needs to be maintained in the oxidized state to retain its CA activity as exposure to thiol reducing agents reduces its CA activity significantly. This likely indicates that disulfide bond formation in Cah3 is essential for its CA activity. Cah3 has three cysteine residues in the mature protein. Site directed mutagenesis experiments can be done to specifically substitute these cysteines in Cah3 to study whether or not cysteines play a role in controlling the CA activity of Cah3. The cytoplasm of cells and the stroma of chloroplasts in general is expected to have a greater reducing environment than that inside the endoplasmic reticulum or thylakoid lumen (Personal communication Dr. Bob Buchanan, University of California, Berkeley, California). The redox environment in the thylakoid lumen can favor the oxidized form of the enzyme which is more active than the reduced form.

All the CA activity measurements and studies related to the partial characterization of the CA activity of Cah3 presented here are based on *in vitro* analyses of the recombinant Cah3 protein and need to be confirmed by *in vivo* analyses. *In vivo* activity measurement of Cah3 is a challenging task because of the presence of multiple isoforms of CAs from the three CA families in different cell compartments in *C. reinhardtii.*

Western blotting results showed that the "new" Cah3 antibody detects Cah3 more specifically than the "old" Cah3 primary antibody that was raised against the MBP-Cah3 fusion protein. The "old" Cah3 antibody cross reacts strongly with two different proteins that had similar molecular weights. These two proteins were not resolved on a 12% or 15% gel as they

migrated together on the gel (Fig. 3.7). Hence, with the "old" Cah3 antibody it was impossible to deduce if *ca-1* and *cia3* mutants lack the Cah3 protein, when protein samples were separated on a 12%-15% SDS-PAGE gel. *cia3* has two point mutations in the region coding for the thylakoid targeting transit peptide of the Cah3 protein. These mutations lead to the mistargeting of the Cah3 protein. *ca-1* has a stop codon leading to a truncated protein, only thirteen amino acids long. In *ca-1*, it is expected that such a highly truncated protein would be degraded. In *cia3*, the mistargeted protein could be either degraded or be present in the stroma of the chloroplast. Western blots with the "new" Cah3 antibody seem to indicate that the protein is missing in *ca-1* and *cia3*. This result needs to be confirmed by the immunolocalization studies. Currently immunolocalization experiments are being done using the "new" Cah3 antibody to determine if the two *cah3* mutants namely *cia3* and *ca-1*, in reality, lack or underexpress the Cah3 protein.

# CHAPTER 4

# IDENTIFICATION OF A NEW β-CA GENE (*Cah6*) AND PARTIAL CHARACTERIZATION OF ITS PROTEIN PRODUCT

## INTRODUCTION

β-CAs were the first carbonic anhydrases to be identified in photosynthetic organisms (Burnell *et al*.,1990; Fawcett *et al*., 1990). The distribution of β-CAs does not appear as widespread as the α-CAs. β-CAs have been found in higher plants, micro-algae, eubacteria (Hewett-Emmett and Tashian,1996), archaebacteria (Smith and Ferry,1999), the fungi *Saccharomyces cerevisiae* and *S. pombe* (Gőtz, *et al*.,1999), *Caenorhabditis elegans* (putative β-CA genes exist in the EST database), *Drosophila melanogaster* (putative β-CA genes exist in the EST database) but not in vertebrates. In $C_3$ plants, an abundant β-CA is localized to the chloroplast stroma and is highly active.

β-CAs have been found both in the cytoplasm ($C_4$ mesophyll cells) and chloroplast ($C_3$ plants) of higher plants and in the symbiotic green alga *Coccomyxa* (Hiltonen *et al*., 1998). In *A. thaliana,* cDNAs encoding cytoplasmic and chloroplastic forms of β-CAs have been described (Fett and Coleman, 1994). At this time sequences encoding at least five β-CA genes from *A. thaliana* are in the EST database but only the stroma and cytoplasmic locations of these proteins have been confirmed. All β-CAs have a histidine and two cysteine residues that act as zinc ligands (Hewett-Emmett and Tashian, 1996). β-CAs are generally more insensitive to inhibition by sulfonamide compounds than the α-CAs (Hewett-Emmett and Tashian, 1996). More information about β-CAs may be found in Chapter 1.

Two mitochondrial β-CAs have been identified previously in the green alga *C. reinhardtii* (Eriksson *et al*., 1996). Mitochondrial β-CAs are induced strongly under low $CO_2$ but

not under high $CO_2$ conditions (Eriksson *et al*., 1998). Here I report the identification a new β-CA gene, *Cah*6, in *C. reinhardtii*, and partial characterization of the recombinant Cah6 protein. The Cah6 protein is targeted to the chloroplast stroma. Cah6 is expressed both under high (5% $CO_2$) and low $CO_2$ (0.035% of $CO_2$ in air) conditions but is slightly upregulated under low $CO_2$ conditions.

**RESULTS**

**1. Screening of the Cosmid and cDNA Library for Genomic and cDNA Clones of *Cah6***

To find new β-CA genes, the mitochondrial β-CA protein (Ca1) sequence of *C. reinhardtii* was used to BLAST the EST database of *C. reinhardtii*. The search yielded many ESTs that were the gene products of the mitochondrial β-CA genes. However, several other ESTs also were found that, while encoding a β-CA, did not belong to either of the two known mitochondrial β-CA genes. These ESTs were from the same gene and were analyzed by CAP (contig assembly program; http://www.infobiogen.fr/services/analyseq/cgi-bin/cap_in.pl) to form a consensus *Cah6* sequence. In order to amplify *Cah6*, several PCR primers were designed based on this *Cah6* contig. The alignment and sequences of all *Cah6* primers used for PCR and sequencing of *Cah6* are given in Appendix 1 of this dissertation.

The PCR primers F4 and R5 (Appendix 1), were used on an indexed cosmid library to isolate a cosmid carrying the *Cah6* gene. After several rounds of screening by PCR, two cosmid clones designated, 72-E-6 and 29-D-12, were isolated. Positive cosmid clones were verified as *Cah6* clones after each round of screening by PCR, followed by sequencing of the PCR product using different *Cah6* primers (Appendix 1). PCR using primers F4 and R5 yielded a product of 2.8 kb when used on the isolated cosmid clones (Fig. 4.1). The same two primers used on the cDNA core library yielded a PCR product of 2.4 kb which is six base pairs short from being a full length Cah6 cDNA clone [excluding the poly A tail; (Fig. 4.2)]. Detailed information on the

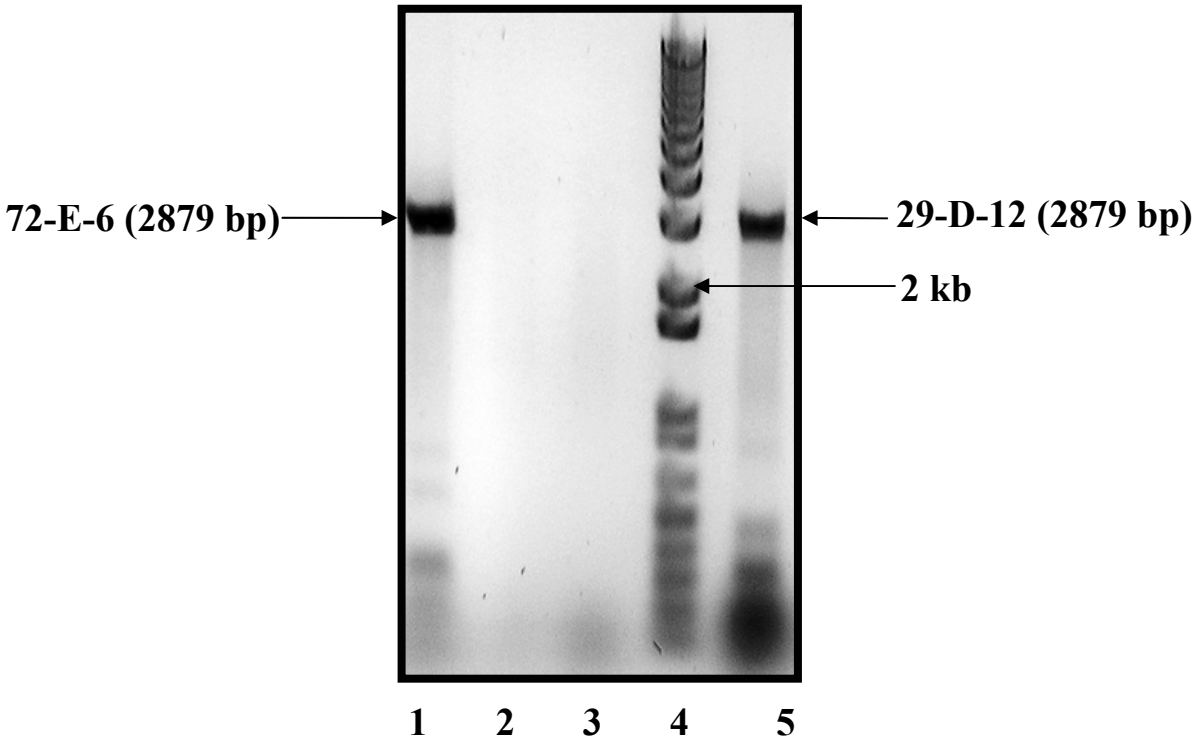



**Figure 4.1 The screening of the cosmid library using the F4 and R5 primers.** Lanes 1 and 5 show the PCR results using the cosmid library. Lanes 2 and 3 show the PCR results in the absence of DNA. Lane 4 shows a 1 kb DNA ladder (NEB). 72-E-6 and 29-D-12 are the two cosmids isolated containing the *Cah6* gene.

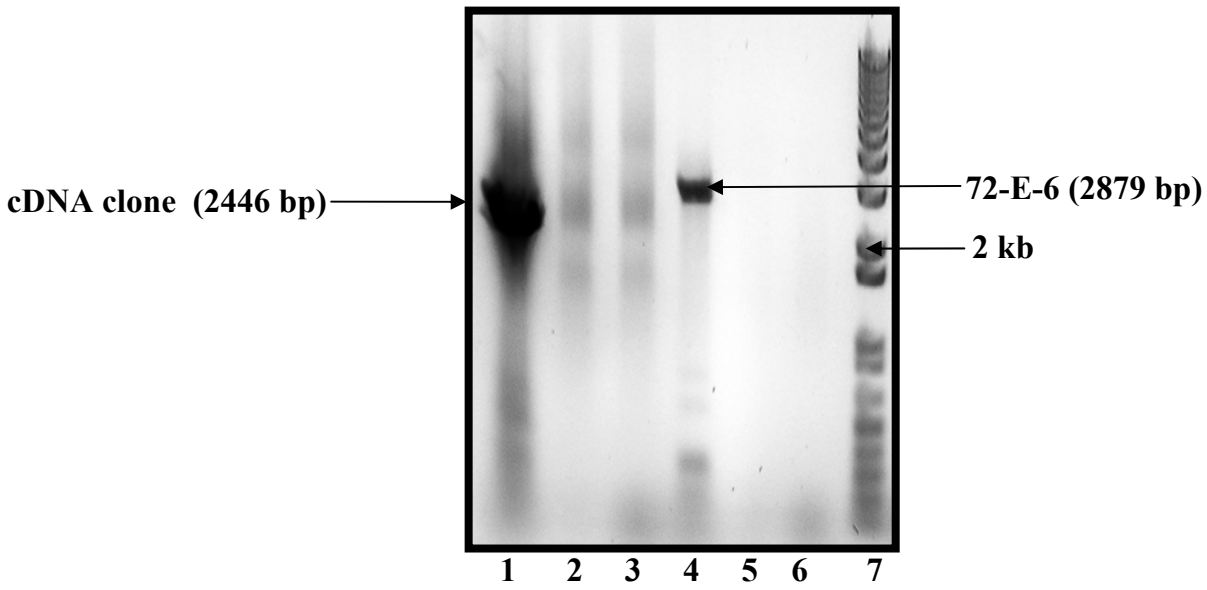



**Figure 4.2 The screening of the cDNA core library for a Cah6 cDNA clone using PCR.** Lane 1 shows the PCR result using the F4 and R5 *Cah6* primers on the cDNA core library. Lanes 2 and 3 show the PCR results using only a single primer F4 and R5, respectively on the cDNA library. Lane 4 shows PCR result using F4 and R5 primers on cosmid 72-E-6. Lanes 5 and 6 represent the PCR products in the absence of DNA. Lane 7 contains a 1 kb DNA ladder (NEB).

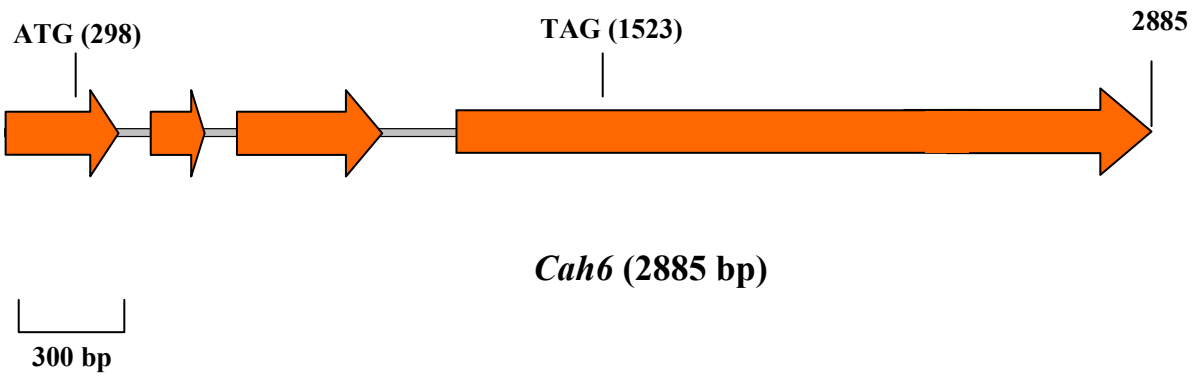



**Figure 4.3 The genomic map of *Cah6*.** The orange block arrows represent the four exons while the lines connecting the block arrows represent the three introns. The start and stop codons are labeled by black lines. The numbers within the parentheses denote the start and stop codon positions on the map in base pairs.

screening of the cosmid library and preparation of a cDNA library can be found in Chapter 2 of this dissertation.

**2. Sequencing and Homology Search**

The cosmid clones 72-E-6 and 29-D-12, and the 2.4 kb cDNA PCR product mentioned above, were sequenced in both directions using primers shown in Appendix 1. The sequencing results were confirmed by the EST and genomic database of *Chlamydomonas*. The *Cah6* gene has four exons and three introns and is 2885 bp long (Fig. 4.3). The size ranges of the exons are from 93 bp to 1652 bp while the introns range from 75 bp to 188 bp long. The genomic map and the full genomic sequence are given in Fig. 4.3 and Fig. 4.4, respectively.

The *Cah6* cDNA is 2452 bp long (Fig. 4.5). It encodes a putative protein of 264 amino acids. It contains a translation start site at nucleotide 309 and a stop site at nucleotide 1101 (Fig. 4.6). It has a very long 3′ UTR, more than 1300 base pairs long. A restriction map of the full length cDNA clone of *Cah6* is shown in Fig. 4.7. Predictions based on various protein prediction programs (SORT P, CHLOR P AND TARGET P) listed under ExPasy tools (http://ca.expasy.org/tools/#translate), indicate that the Cah6 protein is targeted to the chloroplast because it includes a putative chloroplastic transit peptide of 39 amino acids. A protein database search using the Cah6 protein sequence showed that it is similar to β-CAs from *Escherichia coli*, green algae and higher plants with an amino acid identity of 23% to 34%. A multiple sequence alignment of the Cah6 protein with that of β-CAs from other green algae and higher plants shows that it contains the characteristic one histidine and two cysteine residues as zinc coordination residues, seen in the enzymatically active β-CAs (Fig. 4.8).

**3. Cloning of the *Cah6* in an Overexpression Vector**

*Cah6* was cloned into the expression vector pMal to study the properties of Cah6 protein and to raise an antibody against it, if Cah6 were found to be enzymatically active. Two different

CAGATGATACCAGCAACATTAAACACTTATCTTACCACACGCGCGTAGAAGCCATTCAACCTTCTTGTGCAATCCGA
CTCACCTGCCTCCGAGAACACACGCAGCGTCTACCCTCGCACGGCCAGGGGCATTCCTTTCCAAGCCCAGCGGACGG
AAACTGAGCCTGTAGTAGACGAGCAGCAACCATATCTACGCGGTTATCGGTCAGCAATTATACTTTCGTCCCGAGCC
GCGCCCCAGCGGACTCTCGGCGCTCTCTTTGTGGCCCCACAGCCTTAGGCTCACCTAAGCTTCAACC**ATG**gtgagcg
aagaccggcggccgtgccgcggcagtgccacctgaggctcggaactaaacaattcctttccgtcgcagGGATGCGGT
GCCAGCGTGCCTCAGAATGGTGGAGGAGCTCCCGTTACGCGGGTTATGCCCGCGCCAGCACAACCAGTGTCTGAGGC
GCAATCGgtgaggctaatgtggaaattttggtgtcggttcgtccgggatcacgcaggcctgggcgcggcttgagttc
ttgacatgtatccagtaaaagcccatggttgcaagccagtgcggcggcagtggcggtgctgactgatacttggttca
tgttcgtgaactccctctcgcagGCAATCAGCTTCCAACCATCGCGCAGCAACCGCAGCAGCCTTGAAAAGATCAAT
TCGCTCACGGATAGGGCATCGCCTGAGCAGGTGCTGCAGAACCTGCTGGACGGCAACATGCGCTTCCTGGATGGCGC
CGTCGCGCATCCCCACCAGGACTTCAGCCGCGTGCAGGCCATTAAGGCCAAGCAAAAGCCCCTCGCGGCCATCCTGG
GCTGCGCCGACTCTCGCGTGCCTGCGGAAATTGTGTTCGACCAAGGCTTTGGCGACGTGTTCGTGTGCCGTGTCGCC
GGCAACATTGCTACGCCAGAGGAGATCGCCAGTCTGGAGTATGCCGTGCTTGACCTCGGAGTTAAGGTGGTGATGGT
CCTCGGACACACACGCTGCGGAGCCGTGAAGGCTGCACTTTCAGgtggggcggcaagagcgaatttgaagtgcggag
gagggcgtgcacggcgcgtggactggccgagcgccaacgtctcctgccagcgtcccatgcgtagacatctttcgtgc
gggtttgctgacagcgctgattgccaccgccccttacactgcatgccttccctcgctccccgccccctcccccaca
gGCAAGGCGTTCCCCGGCTTCATCGACACGCTGGTGGACCACCTGGACGTCGCCATCAGCCGCGTCAACAGCATGAG
CGCCAAGGCGCACCAGGCCATCAAGGACGGCGACGTGGACATGCTGGACCGCGTGGTGAAGGAGAACGTCAAGTACC
AGGTGCAGCGGTGCCAGCGCTCCGTCATCATCCAGGAGGGGTTGCAGAAGGGGAACCTGCTGCTGGCGGGCGCCGTG
TACGACCTGGACACGGGCAAGGTGCACGTCAGCGTCACCAAGGGCGGCAGCAGCGCCGAG**TAG**CAGTAAGCAGGGTT
GCAGGCAAGGGCTGGGAGACATGCGGTGGCGAGGAAACTGGGATGCAGGCCGTCCACGGTTTGATAAATTATTGTTT
TTTGCATGGTTGTGACATTAGCGAAACACATCATCCTCTCTGACATTGTGCGCTGGCTGAGAGGGTGGCGGAAAGAG
TACTAACACGAGGTCCGCTCGTGATTCTGGAAATGCTGAAGCGTTGCAAGCTCGGAGGCCCACAGCTACAAGGAACG
GATGCATCAAGCTCGGCCGGCCCGTGTAGCCCCGCGCATTTAGGGACTTCATGGCGCAAGGTGTCAGGGCAGCAGCT
GCCGTGTCCGTTTGTCCAGACGTGCGGGTGGCGATGGGTCACAGCGGCAGTAATGTCGTGCCGTGTGGTTTGGCGTG
CCGGTGTTATATTCGTGATCTGCCCGTGTATCGTATGTAGCGCTGGCTTGCAGTGCAACTGGCGCGCTGCGTCCTGC
TGGGCAACGAGTAGGGCTGCGCGCGGCAGGTTTTTGTCCCTGAGCAGGGTGAATGGAAACTCAACTCCTTCATAATA
GGCTCGAAATTCAATAATTGACTGAGCAGCGGTCCGGCAACCGATGCAAAGATGGGCGCTTGTTGCGAGCGCCTGTC
GCCGGATTGTCGGGCAGTTGCGTGTTTCCATGGATGCCCCTGTTTTATGCGGTTGCGGTCTGCCGGTTGAGCCGTGG
ACTAGCCCGCAATATGCAGGCCTGGCATTGTGAACGGGGGTGTCTGTGTAGCCGCACACGCGTGTGCACGCTTTACT
TGAGATTCACGAAGGTAGCTGGTCTACCTGCAACAGTACTGCGGACGATTACGAGGAGGCCTGCAAGTGTGCACTAG
CACTTGCAGCGCAGGCTGGCAGGATTCGTAGCCCTTGCTGGGGCATGATGTTGTTTGTTTTTACTGTCATGAAGCAG
CACCGCAAGAGCTGCATTTGGTCCCTAAGGCAGCCACCTCCTTGTTACTGTCATGGGTTTTGGTGTTGTGTGGAGGT
CCTTTCTCTCATTGCGGCGGTCGGCGACACGACGTACGGCCCGTGACTGTGCGATGGCCAAGGGCGAGCATGCAGAC
TTGAGGGGTCTCCTGAGCGCTTGACAATCCGAGGGATGTATCTGCGAGGGCGAAGAATTCTGATCGTGTCAGACGAA
GACGTTTGCAGCATTGGCGAGCTGGATGGCTGCGGTGTAGCGTGCACAGTGCTGCCCAAGAGGTGTTGGTGTATGTT
TCGGTCGGGCAAGCTTGGGACACCATAACGGTTCAATAAAGCGACTTCCCCTGCATAAGGGGCACGTTGCCACAACT
ACTGGTCGCAGACAATGACTGTAATCTGTACCGCAG

**Figure 4.4 The genomic sequence of *Cah6*.** The genomic sequence is 2885 bp long. The exon sequences are in uppercase and the intron sequences are in lower case. The start and stop codons are shown in bold red.

CAGATGATACCAGCAACATTAAACACTTATCTTACCACACGCGCGTAGAAGCCATTCAACCTTCTTGTGCAATCCGA
CTCACCTGCCTCCGAGAACACACGCAGCGTCTACCCTCGCACGGCCAGGGGCATTCCTTTCCAAGCCCAGCGGACGG
AAACTGAGCCTGTAGTAGACGAGCAGCAACCATATCTACGCGGTTATCGGTCAGCAATTATACTTTCGTCCCGAGCC
GCGCCCCAGCGGACTCTCGGCGCTCTCTTTGTGGCCCCACAGCCTTAGGCTCACCTAAGCTTCAACC**ATG**GGATGCG
GTGCCAGCGTGCCTCAGAATGGTGGAGGAGCTCCCGTTACGCGGGTTATGCCCGCGCCAGCACAACCAGTGTCTGAG
GCGCAATCGGCAATCAGCTTCCAACCATCGCGCAGCAACCGCAGCAGCCTTGAAAAGATCAATTCGCTCACGGATAG
GGCATCGCCTGAGCAGGTGCTGCAGAACCTGCTGGACGGCAACATGCGCTTCCTGGATGGCGCCGTCGCGCATCCCC
ACCAGGACTTCAGCCGCGTGCAGGCCATTAAGGCCAAGCAAAAGCCCCTCGCGGCCATCCTGGGCTGCGCCGACTCT
CGCGTGCCTGCGGAAATTGTGTTCGACCAAGGCTTTGGCGACGTGTTCGTGTGCCGTGTCGCCGGCAACATTGCTAC
GCCAGAGGAGATCGCCAGTCTGGAGTATGCCGTGCTTGACCTCGGAGTTAAGGTGGTGATGGTCCTCGGACACACAC
GCTGCGGAGCCGTGAAGGCTGCACTTTCAGGCAAGGCGTTCCCCGGCTTCATCGACACGCTGGTGGACCACCTGGAC
GTCGCCATCAGCCGCGTCAACAGCATGAGCGCCAAGGCGCACCAGGCCATCAAGGACGGCGACGTGGACATGCTGGA
CCGCGTGGTGAAGGAGAACGTCAAGTACCAGGTGCAGCGGTGCCAGCGCTCCGTCATTCATCCAGGAGGGGTTGCAG
AAGGGGAACCTGCTGCTGGCGGGCGCCGTGTACGACCGGACACGGGCAAGGTGCACGTCAGCGTCACCAAGGGCGGC
AGCAGCGCCGAG**TAG**CAGTAAGCAGGGTTGCAGGCAAGGGCTGGGAGACATGCGGTGGCGAGGAAACTGGGATGCAG
GCCGTCCACGGTTTGATAAATTATTGTTTTTTGCATGGTTGTGACATTAGCGAAACACATCATCCTCTCTGACATTG
TGCGCTGGCTGAGAGGGTGGCGGAAAGAGTACTAACACGAGGTCCGCTCGTGATTCTGGAAATGCTGAAGCGTTGCA
AGCTCGGAGGCCCACAGCTACAAGGAACGGATGCATCAAGCTCGGCCGGCCCGTGTAGCCCCGCGCATTTAGGGACT
TCATGGCGCAAGGTGTCAGGGCAGCAGCTGCCGTGTCCGTTTGTCCAGACGTGCGGGTGGCGATGGGTCACAGCGGC
AGTAATGTCGTGCCGTGTGGTTTGGCGTGCCGGTGTTATATTCGTGATCTGCCCGTGTATCGTATGTAGCGCTGGCT
TGCAGTGCAACTGGCGCGCTGCGTCCTGCTGGGCAACGAGTAGGGCTGCGCGCGGCAGGTTTTTGTCCCTGAGCAGG
GTGAATGGAAACTCAACTCCTTCATAATAGGCTCGAAATTCAATAATTGACTGAGCAGCGGTCCGGCAACCGATGCA
AAGATGGGCGCTTGTTGCGAGCGCCTGTCGCCGGATTGTCGGGCAGTTGCGTGTTTCCATGGATGCCCCTGTTTTAT
GCGGTTGCGGTCTGCCGGTTGAACCGTGGACTAGCCCGCAATATGCAGGCCTGGCATTGTGAACGGGGGTGTCTGTG
TAGCCGCACACGCGTGTGCACGCTTTACTTGAGATTCACGAAGGTAGCTGGTCTACCTGCAACAGTACTGCGGACGA
TTACGAGGAGGCCTGCAAGTGTGCACTAGCACTTGCAGCGCAGGCTGGCAGGATTCGTAGCCCTTGCTGGGGCATGA
TGTTGTTTGTTTTTACTGTCATGAAGCAGCACCGCAAGAGCTGCATTTGGTCCCTAAGGCAGCCACCTCCTTGTTAC
TGTCATGGGTTTTGGTGTTGTGTGGAGGTCCTTTCTCTCATTGCGGCGGTCGGCGACACGACGTACGGCCCGTGACT
GTGCGATGGCCAAGGGCGAGCATGCAGACTTGAGGGGTCTCCTGAGCGCTTGACAATCCGAGGGATGTATCTGCGAG
GGCGAAGAATTCTGATCGTGTCAGACGAAGACGTTTGCAGCATTGGCGAGCTGGATGGCTGCGGTGTAGCGTGCACA
GTGCTGCCCAAAAGGTGTTGGTGTATGTTTCGGTCGGGCAAGCTTGGGACACCATAACGGTTCAATAAAGCGACTTC
CCCTGCATAAGGGGCACGTTGCCACAACTACTGGTCGCAGACAATGAC**TGTAA**TCTGTACCGCAG

**Figure 4.5 The full length cDNA sequence of *Cah6*.** The cDNA sequence is 2452 bp in length. The start and stop codons are shown in bold red. The polyadenylation signal is shown in bold blue.

```
 309 atgggatgcggtgccagcgtgcctcagaatggtggaggagctccc
     M  G  C  G  A  S  V  P  Q  N  G  G  G  A  P
 354 gttacgcgggttatgcccgcgccagcacaaccagtgtctgaggcg
     V  T  R  V  M  P  A  P  A  Q  P  V  S  E  A
 399 caatcggcaatcagcttccaaccatcgcgcagcaaccgcagcagc
     Q  S  A  I  S  F  Q  P  S  R  S  N  R  S  S
 444 cttgaaaagatcaattcgctcacggatagggcatcgcctgagcag
     L  E  K  I  N  S  L  T  D  R  A  S  P  E  Q
 489 gtgctgcagaacctgctggacggcaacatgcgcttcctggatggc
     V  L  Q  N  L  L  D  G  N  M  R  F  L  D  G
 534 gccgtcgcgcatccccaccaggacttcagccgcgtgcaggccatt
     A  V  A  H  P  H  Q  D  F  S  R  V  Q  A  I
 579 aaggccaagcaaaagcccctcgcggccatcctgggctgcgccgac
     K  A  K  Q  K  P  L  A  A  I  L  G  C  A  D
 624 tctcgcgtgcctgcggaaattgtgttcgaccaaggctttggcgac
     S  R  V  P  A  E  I  V  F  D  Q  G  F  G  D
 669 gtgttcgtgtgccgtgtcgccggcaacattgctacgccagaggag
     V  F  V  C  R  V  A  G  N  I  A  T  P  E  E
 714 atcgccagtctggagtatgccgtgcttgacctcggagttaaggtg
     I  A  S  L  E  Y  A  V  L  D  L  G  V  K  V
 759 gtgatggtcctcggacacacacgctgcggagccgtgaaggctgca
     V  M  V  L  G  H  T  R  C  G  A  V  K  A  A
 804 ctttcaggcaaggcgttccccggcttcatcgacacgctggtggac
     L  S  G  K  A  F  P  G  F  I  D  T  L  V  D
 849 cacctggacgtcgccatcagccgcgtcaacagcatgagcgccaag
     H  L  D  V  A  I  S  R  V  N  S  M  S  A  K
 894 gcgcaccaggccatcaaggacggcgacgtggacatgctggaccgc
     A  H  Q  A  I  K  D  G  D  V  D  M  L  D  R
 939 gtggtgaaggagaacgtcaagtaccaggtgcagcggtgccagcgc
     V  V  K  E  N  V  K  Y  Q  V  Q  R  C  Q  R
 984 tccgtcatcatccaggaggggttgcagaaggggaacctgctgctg
     S  V  I  I  Q  E  G  L  Q  K  G  N  L  L  L
1029 gcgggcgccgtgtacgacctggacacgggcaaggtgcacgtcagc
     A  G  A  V  Y  D  L  D  T  G  K  V  H  V  S
1074 gtcaccaagggcggcagcagcgccgagtag 1103
     V  T  K  G  G  S  S  A  E  *
```

**Figure 4.6 The amino acid translation of the *Cah6* cDNA clone.** This protein shows an open reading frame of 264 amino acids. The start and stop codons are in bold red. The putative transit peptide sequence is shown in bold green.

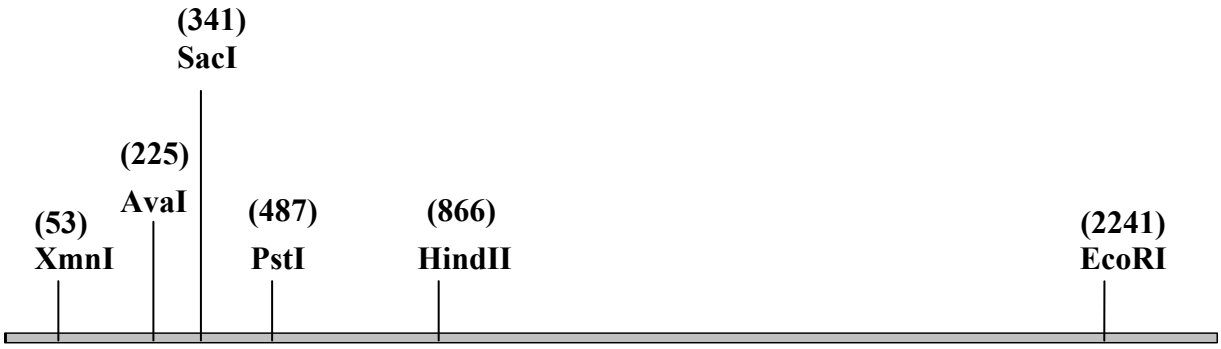



**Figure 4.7 The restriction map of the Cah6 cDNA.** The restriction enzyme sites are indicated by black lines. The nucleotide numbers within the parentheses indicate the cleavage sites of the enzymes.

```
Cah6            ----------VSEAQSAISFQPSRSNRS----SLEKINSLTDRAS-PEQVLQNLLDGNM
Ca1             ----------ASAVNKGCSCRCGRVACMGACMPMRHLHAHPNPPSDPDQALEYLREGNK
Coccomyxa       ----------------------------------------MSAKDTADLSPLLEANR
Spinach         TLKEDMAYEEAIAALKKLLSEKGELENEAASKVAQITSELADGGTPSASYPVQRIKEGFI
                                                         ..    :. : :.

Cah6            RFLDG--AVAHPHQDFSRVQAIKAKQKPLAAILGCADSRVPAEIVFDQGFGDVFVCRVAG
Ca1             RFVNNKPHDSHPTRNLDRVKATAAGQKPFAAFLSCADSRVPVEIIFDQGFGDVFVTRVAG
Coccomyxa       KWADE-----CAAKDSTYFSKVAGSQAPEYLYIGCADSRVSPAQLFNMAPGEVFVQRNVG
Spinach         KFKKE-----KYEKNPALYGELSKGQAPKFMVFACSDSRVCPSHVLDFQPGEAFMVRNIA
                :: .         ::          * *    :.*:****    :::   *:.*: *  .

Cah6            NIATP-------EEIASLEYAVLDLGVKVVMVLGHTRCGAVKAALSGKAFPGFIDTLVDH
Ca1             NIVTN-------EITASLEFGTAVLGSKVLMVLGHSACGAVAATMNGAAVPGVISSLYYS
Coccomyxa       NLVSNKD----LNCMSCLEYTVDHLKIKHILVCGHYNCGACKAGLVWHPKTAGVTNLWIS
Spinach         NMVPVFDKDKYAGVGAAIEYAVLHLKVENIVVIGHSACGGIKGLMSFPDAGPTTTDFIED
                *:..           :.:*: .  *  : ::* **  **.  . :           :

Cah6            LDVAISRVNSMSAKAHQAIKDGDVDMLDRVVKENVKYQVQRCQRSVIIQEGLQKGN-LLL
Ca1             ISPACKKA----------QAGDVDG---AIAENVKVQMEQLKVSPVLQGLVKEGK-LKI
Coccomyxa       DVREVRDKN-----AAKLHGLSADDAWDKMVELNVEAQVFNVCASPIVQAAWARGQPLSV
Spinach         WVKICLPAK--HKVLAEHGNATFAEQCTHCEKEAVNVSLGNLLTYPFVRDGLVKKT-LAL
                                :        *: .: .     .::    .   * :

Cah6            AGAVYDLDTGKVHVSVTKGGSSAE------------------- 264
Ca1             VGGVYDLATGKVTEIA--------------------------- 267
Coccomyxa       HGIVYTPGTGLVKELIKPITGMEDAGALLRADLKQHCFFSESLA 227
Spinach         QGGYYDFVNGSFELWGLEYGLSPSQSV---------------- 319
                 *  *   .* .
```

**Figure 4.8 The alignment of the Cah6 protein sequence with those of other well characterized β-CAs.** Ca1 represents the mitochondrial β-CA. Mitochondrial β-Ca1 and β-Ca2 (sequence not shown in the alignment) are almost identical in amino acid sequence and have only one amino acid difference in their sequences. *Coccomyxa* represent the cytosolic β-CA from the symbiotic alga *Coccomyxa.* The leader sequence of spinach which is 98 bp long is not shown here. Active site residues are shown in bold red. ***** represent a completely conserved amino acid, **:** represent conserved amino acid substitutions, and **.** represent semi conserved amino acid substitutions.

*MalE-Cah6* recombinant overexpression constructs were generated. One contained the cDNA sequence coding for the full length open reading frame (ORF) of the Cah6 protein and the other contained the cDNA sequence that would code for the mature Cah6 protein predicted by the web protein prediction programs. The SP2 + R4H primers and the MP + R4H primers were used to amplify the cDNA coding for the full length and putative mature Cah6 proteins, respectively (Fig. 4.9).

Amplified cDNAs were purified from the gel and cloned into the overexpression vector pMal-c2x vector (Fig. 4.10) to generate two different types of recombinant constructs as described in Chapter 2. An in-frame insertion of *Cah6* with the sequence of MBP in the recombinant clone was verified by restriction enzyme digestion analyses (Fig. 4.11) and DNA sequencing of the inserted Cah6 cDNA sequence. The latter was done to check for any missense or null mutation that could have been introduced by the DNA polymerase during PCR. Two clones were selected out of the twenty five recombinant clones of each type. Clone B48 had a cDNA insert coding for the full length Cah6 protein while clone B3 had one coding for the mature Cah6 protein.

**4. Overexpression of the MBP-Cah6**

*Escherichia coli* cells harboring the recombinant B48 and B3 recombinant constructs were induced with 1 mM IPTG for 2 hours at 37°C to overexpress the MBP-Cah6 fusion protein (Chapter 2). Equal amounts of proteins from induced and uninduced cells were loaded on a 12% SDS-polyacrylamide gel and subjected to electrophoresis (Fig. 4.12). The overexpressed recombinant Cah6 fusion protein was 15% of the total *E. coli* cell protein.

**5. Purification and Activity Assays of the Recombinant Cah6**

The crude cell extracts of B48 and B3 clones were used for CA activity assays. CA activity was detected in the cell extracts of the induced B48 clone but not in that of the B3 clone.

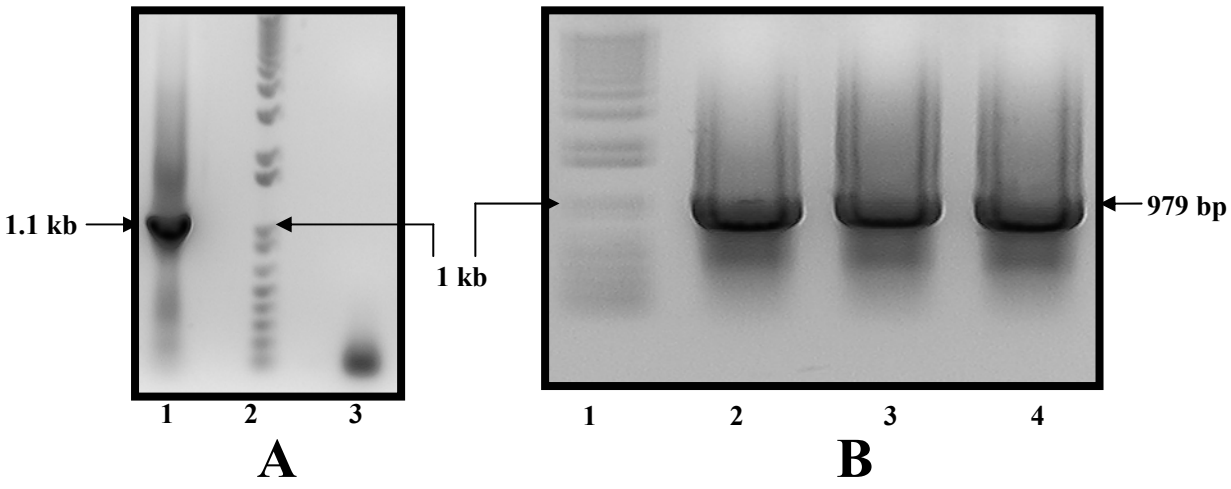



**Figure 4.9 PCR for the construction of the pMal-*Cah6* overexpression plasmid.** A. Lane 1 shows the PCR result using SP2 and R4H primers on the cDNA library. Lane 2 shows a 1 kb DNA ladder. Lane 3 shows the PCR result using the same set of primers in the absence of DNA. B. Lane 1 shows a 1 kb DNA ladder (NEB). Lanes 2, 3 and 4 (replicates) show the PCR results using the MP and R4H primers on the cDNA library.

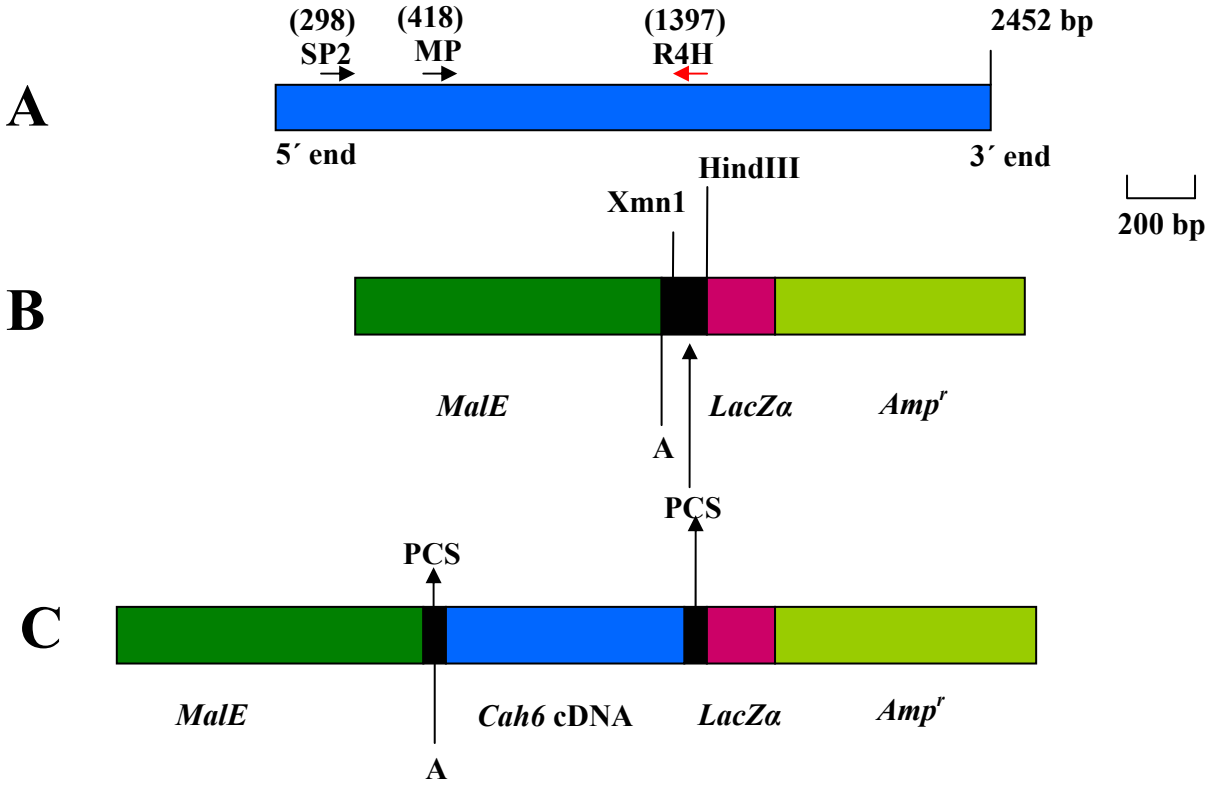



**Figure 4.10 A schematic diagram of the recombinant pMal-Cah6 expression construct.** A. A schematic figure showing the alignment of primers used to amplify the Cah6 cDNA. R4H primer has a HindIII site incorporated at the 5′end. The numbers on the top denote the primer position in base pairs. B. Part of the pMal vector showing the polylinker cloning site (PCS), β-galactosidase (*LacZα*) and β-lactamase (*Amp*$^r$) genes, XmnI and HindIII recognition sites. "A" denotes the Factor Xa cleavage site. *MalE* codes for the maltose binding protein (MBP). C. A schematic figure of the *Cah6* recombinant construct.

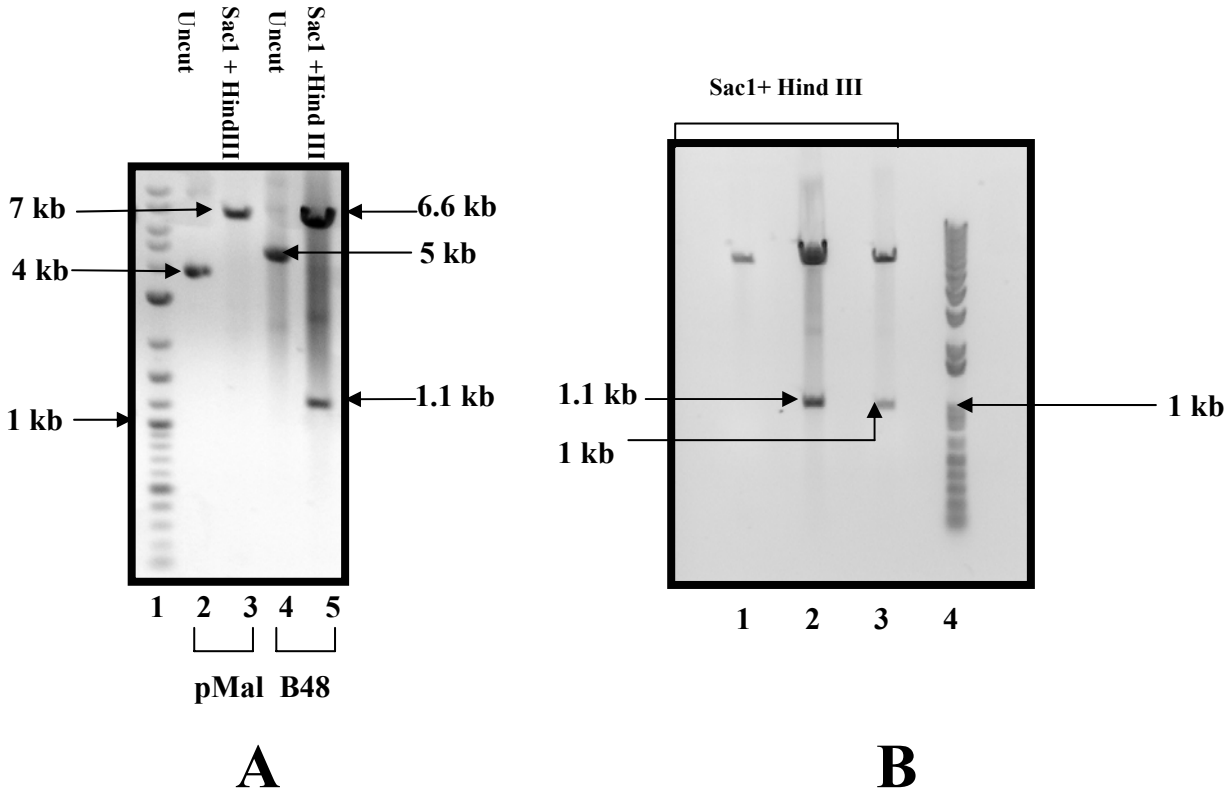



**Figure 4.11 The restriction enzyme digestions of constructs B48, B3 and pMal vector to verify the cloning of Cah6.** A. Lane 1 contains a 2-log DNA ladder (NEB). Lane 2 contains uncut pMal. Lane 3 contains double digested pMal. Lane 4 contains uncut B48. Lane 5 contains double digested B48. B. Lanes 1, 2 and 3 contain double digested pMal, B48 and B3 plasmids, respectively. Lane 4 contains a 1 kb DNA ladder (NEB).

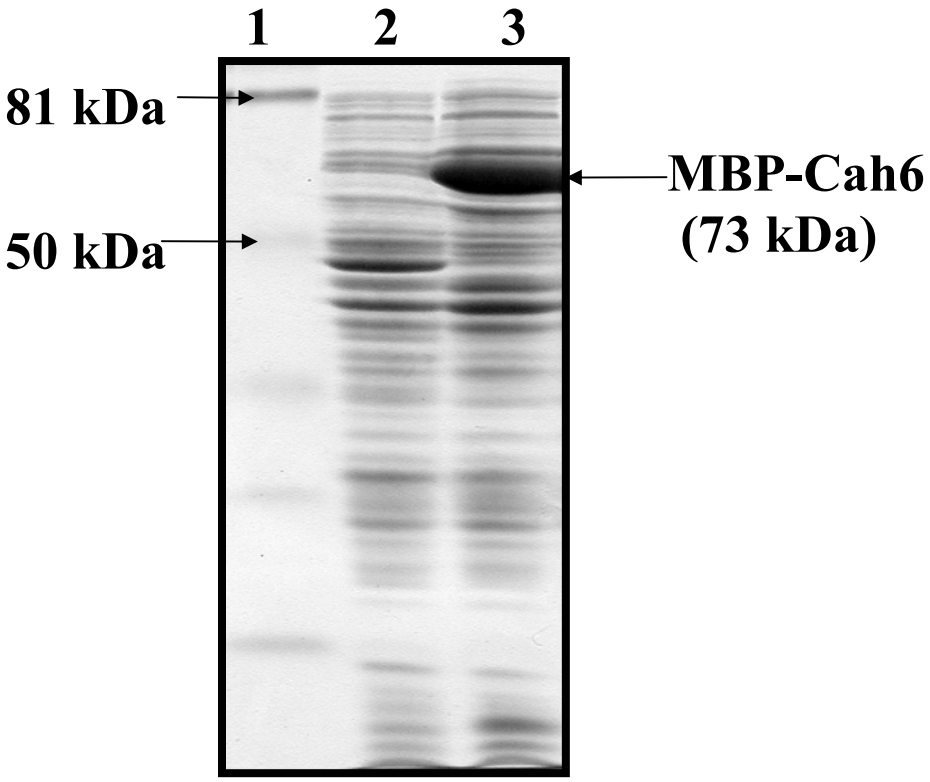



**Fig 4.12 A 12% SDS-polyacrylamide gel showing the overexpression of the recombinant MBP-Cah6.** Lane 1 represents the prestained molecular weight markers. Lanes 2 and 3 represent 30 μg of proteins from uninduced and induced *Escherichia coli* cells, respectively. The B48 clone was used for overexpression.

The B48 clone was selected for Cah6 purification. This clone had the entire ORF (open reading frame) of Cah6. The overexpressed recombinant Cah6 protein was purified by affinity chromatography using amylose resin according to a protocol given in the NEB technical catalogue. Purified recombinant Cah6 was further concentrated by using a 100 kDa cut-off centricon column.

At each step of purification the CA activity in the sample was assayed to check the purity of the Cah6 sample (Table 4.1). The recombinant Cah6 protein was found to have a specific activity of 400 WAU/mg. This calculation of specific activity was based on the total amount of recombinant fusion protein in the sample. CA activity assays were done using the method of Wilbur and Anderson (Wilbur and Anderson, 1948). CA activity was not detected in the extracts from uninduced cells containing the B48 clone or *E. coli* cells containing only the pMal vector.

The fusion protein purified from the B48 clone was cleaved by the protease Factor Xa for four hours at 23°C to separate Cah6 from the MBP. Factor Xa cleaved the fusion protein to yield 42 kDa MBP and 31 kDa Cah6 protein. It also nonspecifically cleaved the 31 kDa Cah6 protein band to yield a 28 kDa fragment. Use of common protease inhibitors like leupeptine did not prevent this nonspecific cleavage of Cah6. Purification and cleavage of the fusion protein was confirmed by performing SDS-PAGE (Fig. 4.13). There was no significant difference between the CA activities of the Factor Xa cleaved MBP-Cah6 fusion protein and uncut fusion protein. The 31 kDa Cah6 protein band was excised from the gel to be used as an antigen for production of polyclonal Cah6 primary antibodies.

**6. Western Blotting and Northern Blotting Analyses of the Cah6 Expression**

To test the specificity of the Cah6 antibody, Factor Xa (protease) cleaved MBP-Cah6 and purified MBP-Cah6 fusion proteins were separated by 12% SDS-PAGE and probed with the Cah6 antibody (Fig. 4.14). The antibody did not react with the MBP when induced *E. coli* cells

**Table 4.1** The purification of the chimeric MBP-Cah6 from 2 liters of *E. coli* culture

| STEP | TOTAL ACTIVITY (WAU)[a] | PROTEIN (mg)[b] | SPECIFIC ACTIVITY (WAU/mg) | RECOVERY (%) | PURIFICATION FOLD |
|---|---|---|---|---|---|
| Sonicated cells | 1200 | 8000 | 0.15 | 100 | 1 |
| Amylose column | 800 | 3 | 300 | 67 | 2000 |
| Centricon column | 400 | 1 | 400 | 33 | 2567 |

[a] one WAU = $(t_0-t)/ t$ where $t_0$ is the time for uncatalyzed reaction and t is the time for the enzyme catalyzed reaction; [b] determined by Lowry's protein assay (Lowry *et al*., 1951)

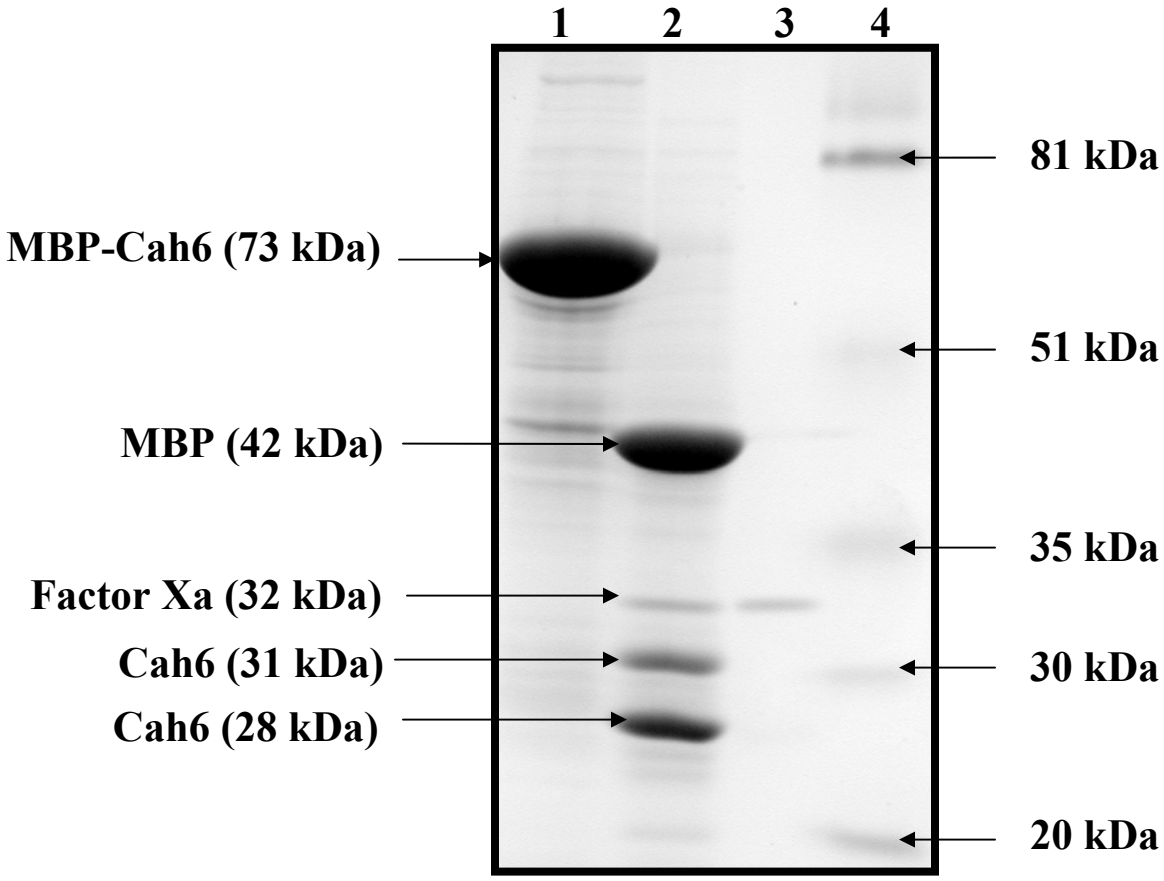



**Figure 4.13 A 12% SDS-polyacrylamide gel showing the uncut and the Factor Xa cut purified MBP-Cah6 protein.** Lane 1 contains 55 µg of uncut purified recombinant protein. Lane 2 contains 55 µg of purified recombinant protein cut by 1 µg of Factor Xa. Lane 3 contains 1 µg of Factor Xa. Lane 4 contains the prestained molecular weight markers. Protein was purified from the B48 clone.

containing only the pMal-c2x vector or Factor Xa cleaved purified MBP-Cah6 fusion protein were used.

Proteins extracted from high $CO_2$ and air adapted D66 cells were separated by electrophoresis and probed with the Cah6 and mitochondrial β-CA primary antibodies (Fig. 4.15). The Cah6 antibody detected the Cah6 protein (28 kDa band) in both the high $CO_2$ and air adapted cells. The air acclimated cells showed slight upregulation of the Cah6 protein compared to that of the high $CO_2$ adapted cells (Fig. 4.15A). The mitochondrial β-CA antibody picked up the 22 kDa mitochondrial β-CA protein in the air acclimated cells but not in the high $CO_2$ grown cells (Fig. 4.15B) in agreement with earlier observations (Eriksson *et al*., 1998).

The primers X-9 and R5 (Appendix 1) were used to amplify an 826 bp PCR product from the cDNA core library (Fig. 4.16). This 826 bp PCR product is contained within the 3′ UTR of *Cah6* and was used as a probe for Northern blot analyses using mRNA extracted from high and low $CO_2$ adapted *Chlamydomonas* (strain D66) cells grown in minimal medium (Fig. 4.16). Twenty μg of RNA from both cell types were loaded on the RNA gel. Northern blotting shows that Cah6 is expressed under both low and high $CO_2$ conditions, but is slightly upregulated under air levels of $CO_2$ cells.

**7. Immunolocalization of Cah6**

Air adapted D66 cells grown in minimal medium were used for immunolocalization of Cah6 following the protocol described in Chapter 2. *Chlamydomonas* cell sections were probed with the Cah6 antibody or the preimmune serum and observed under a transmission electron microscope (Fig. 4.17, Fig. 4.18 and Fig. 4.19). Immunogold densities in different cell compartments are given in Table 4.2. Immunogold densities in different cell compartments in sections were calculated by dividing the number of immunogold particles in a particular cell organelle with the area of that cell organelle. Immunolocalization results demonstrated that

Cah6 is located in the stroma of the chloroplast and is four fold more abundant in the area around the pyrenoid (particularly in the starch sheath of the pyrenoid; Fig. 4.18) compared to the other areas of the stroma.

**8. Partial Characterization of the *Cah6* Activity**

The effects of sulfonamide and anion inhibitors on the CA activity of recombinant Cah6 were studied. Table 4.3 shows the inhibition of recombinant Cah6 by sulfonamides and anions. Generally, all β-CAs are less sensitive to sulfonamide inhibition and slightly more inhibited by anions like $ClO_4^-$, $I^-$, $NO_2^-$, $NO_3^-$, $N_3^-$ etc than are α-CAs (Johansson and Forsman, 1993). Cah6 is no exception. Cah6 was comparatively less inhibited by the sulfonamides and more inhibited by the anions azide and cyanide than bovine CAII which is an α-CA. The effects of different temperatures on the CA activity of the recombinant Cah6 were also studied. The thermostability profile of Cah6 is shown in Fig. 4.20. The enzyme lost more than 50% of its activity above 43°C and completely lost its activity at 50°C. The optimum Cah6 activity was obtained at 33°C. Experiments also were performed with different thiol reducing agents to test the sensitivity of recombinant Cah6 to oxidation. β-CAs localized to higher plant chloroplasts have been reported to be sensitive to oxidation and therefore are dependent on a reducing environment to retain catalytic activity (Tobin, 1970; Atkins *et al*., 1972; Cybulsky *et al*., 1979; Johansson and Forsman, 1993). For example, oxidized pea CA requires a reducing agent for maximal activation of the enzyme (Johansson and Forsman, 1993). In contrast, the activity of Cah6 was found to be independent of a reducing agent. The effects of 10 mM 2-mercaptoethanol, cysteine and dithiothreitol on the Cah6 activity are given in Table 4.5.

**9. Application of RNA Interference (RNAi) to Study the Functional Role of Cah6**

I used RNAi to generate *Cah6* mutants that would lack or underexpress the Cah6 protein. Such mutants would help to probe the functional role played by *Cah6* in CCM and

photosynthesis. Two different RNAi constructs of *Cah6* were generated by cloning reverse complementary sequences from the 5′ and the 3′ region of Cah6 in the pSL72 vector. The 5′ and the 3′ end regions of *Cah6* that were least similar in DNA sequence to those of the mitochondrial β-CAs (*Ca1* and *Ca2*) were selected based on comparative DNA sequence analyses of the *Ca1*, *Ca2* and *Cah6* genes. This was done to specifically underexpress *Cah6* gene and not the other two mitochondrial β-CA genes, which have DNA sequences very similar to that of Cah6 in the coding regions of the gene. RNAi primers (Appendix 2) were designed to amplify reverse complementary fragments of Cah6 from the cDNA core library and cosmid 72-E-6. The alignment of these primers and the PCR results obtained using these primers are shown in Fig. 4.21 and Fig. 4.22, respectively. The 3′ end of the RNAi construct had reverse complementary sequences from a selected region in the 3′ UTR of the *Cah6* cDNA (Fig. 4.23). The 5′ end RNAi construct contained a genomic sequence from the 5′ end of the *Cah6* gene and the corresponding cDNA sequence cloned into the 5′ and 3′ end of the cytochrome $c_6$ intron, respectively, in the pSL72 vector (Fig. 4.23). The genomic fragment had two introns of *Cah6*.

Six 5′ RNAi (5A, 7A, 8A, 10A, 11A and 13A) and five 3′ RNAi clones (1B, 7B, 8B, 10B and 13B) were selected. Double restriction enzyme digestions of these clones with AvrII and XbaI led to the identification of the positive 5′ and 3′ RNAi clones (Fig 4.24). Five 5′ (5A, 7A, 8A, 10A and 11A) and two 3′ RNAi positive clones (1B and 10B) were identified (Fig 4.24). 5′ RNAi clone 11A and 3′ RNAi clone 10B were selected as sources of RNAi plasmids for the transformation of wild type *C. reinhardtii* cells.

Transformation was done using the electroporation method described in Chapter 2. Both linear and circular forms of DNA from the RNAi clones were used for transformation. DNA was linearized by restriction digestion with Kpn1. Transformants were plated on TAP-paromomycin plates under diffuse light.

Three hundred 5′ RNAi and 600 hundred 3′ RNAi paromomycin resistant colonies were picked and regrown on fresh TAP-paromomycin plates. After one week, these colonies were screened on high and low $CO_2$ conditions on MIN plates. At this point there is no difference in the phenotype of the selected transformants under high and and low $CO_2$ conditions compared to that of the wild type *C. reinhardtii* cells. Screening of these transformants by Western blotting using the Cah6 antibody is being done currently. At this point, 5-6 air acclimated transformants of both types have been screened using Western blotting. None of them show any visible difference in the expression of Cah6 protein in the cells compared to that of in the wild type cells.

**DISCUSSION**

Here I report the identification of a nuclear gene encoding a novel chloroplastic β-CA, *Cah6*, and the partial characterization of the Cah6 protein in *C. reinhardtii*. β-CAs were the first CAs recognized in photosynthetic organisms (Burnell *et al*., 1990; Fawcett *et al.*, 1990) but later have been identified in eubacteria, cyanobacteria, yeast, micro-algae and higher plants. They play important roles in these organisms. For example, β-CA has been reported to play an important role in the cyanobacterial CCM (Price and Badger, 1989b). In these organisms a carboxysomal β-CA, coded by *icfA* gene, converts $HCO_3^-$ to $CO_2$ (Badger and Price, 1992). This increases the concentration of $CO_2$ at the site of Rubisco, ensuring efficient $CO_2$ fixation.

Alternatively, it has been suggested that the mitochondrial CAs (Ca1 and Ca2) of *Chlamydomonas reinhardtii* may play roles in recycling both respiratory and photorespiratory $CO_2$ by converting it to $HCO_3^-$ in the mitochondrial matrix (Raven, 2001). This $HCO_3^-$ then would leak back into the cytosol where it would be available for transport into the chloroplast stroma. This model assumes the absence of CA from the cytosol. Recently, it has been shown that the expression of mitochondrial CAs (Ca1 and Ca2) decreases when the external $NH_4^+$ concentration decreases, to the point of being undetectable when the supply of $NH_4^+$ restricts the

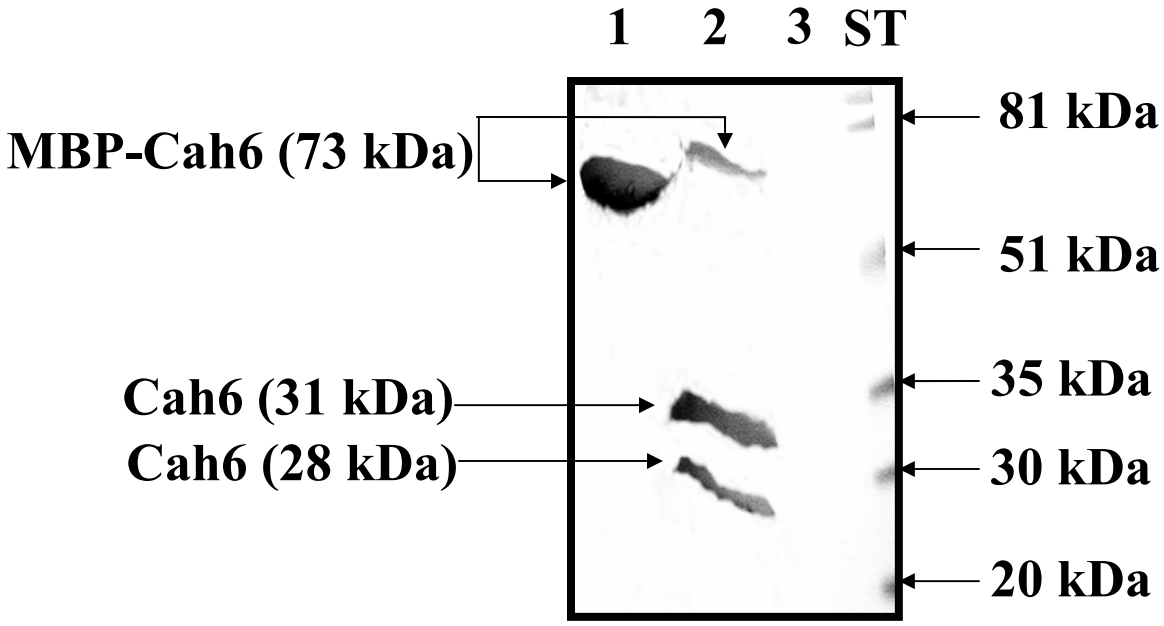



**Figure 4.14 A Western blot probed by the Cah6 antibody using the purified overexpressed MBP-Cah6 protein.** Lane 1 contains 20 μg of purified uncut fusion protein. Lane 2 contains 20 μg of fusion protein cleaved by 1 μg of Factor Xa. Lane 3 contains 1 μg of Factor Xa. ST represents prestained low molecular weight markers.

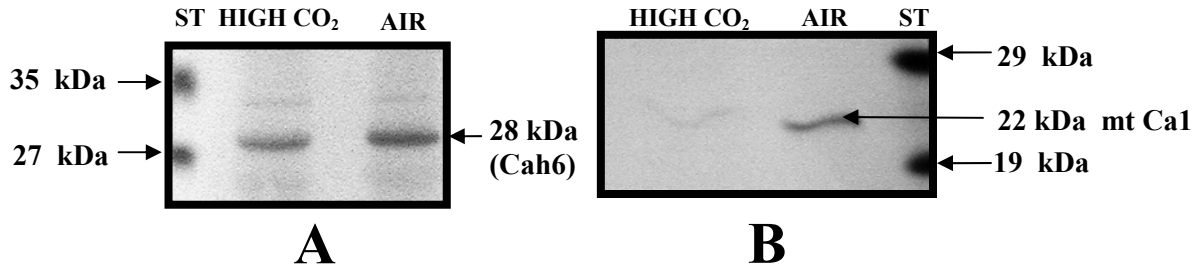



**Figure 4.15 Western blots probed by the Cah6 and mitochondrial β-CA antibodies using the wild type *Chlamydomonas* cells.** A. A Western blot probed by Cah6 antibody using high and low $CO_2$ adapted wild type *Chlamydomonas* (strain D66) grown in minimal medium cells. B. A Western blot probed by mitochondrial β-CA antibody using high (5%) and low $CO_2$ (0.03%) adapted wild type *Chlamydomonas* (strain D66) cells grown in minimal medium.

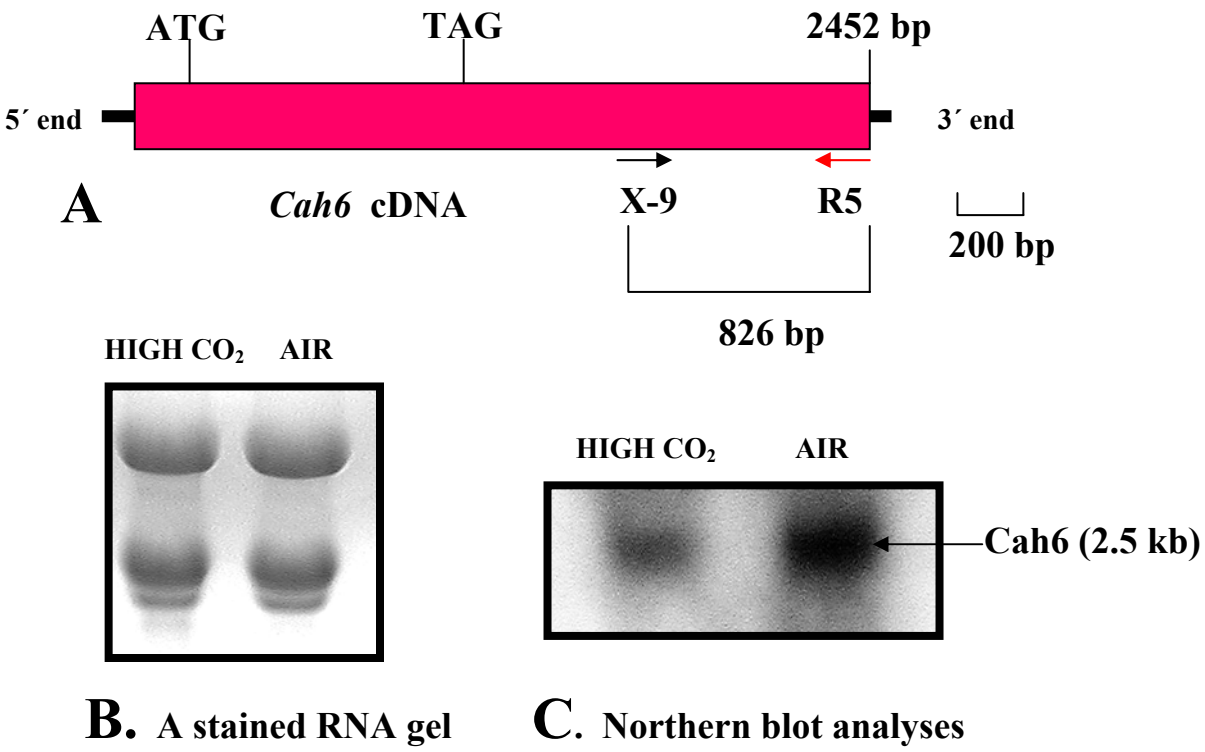



**Fig 4.16 The Northern blot analyses of Cah6 expression.** A. A schematic figure showing the alignment of primers (X-9 and R-5) used for making the probe for Northern blot analyses of Cah6 expression. B. A stained RNA gel showing RNA extracted from high (5%) and low $CO_2$ (0.03%) adapted wild type D66 cells grown in minimal medium. Each lane contains 20 µg of RNA. C. Northern blot result using the 826 bp PCR product as a probe.

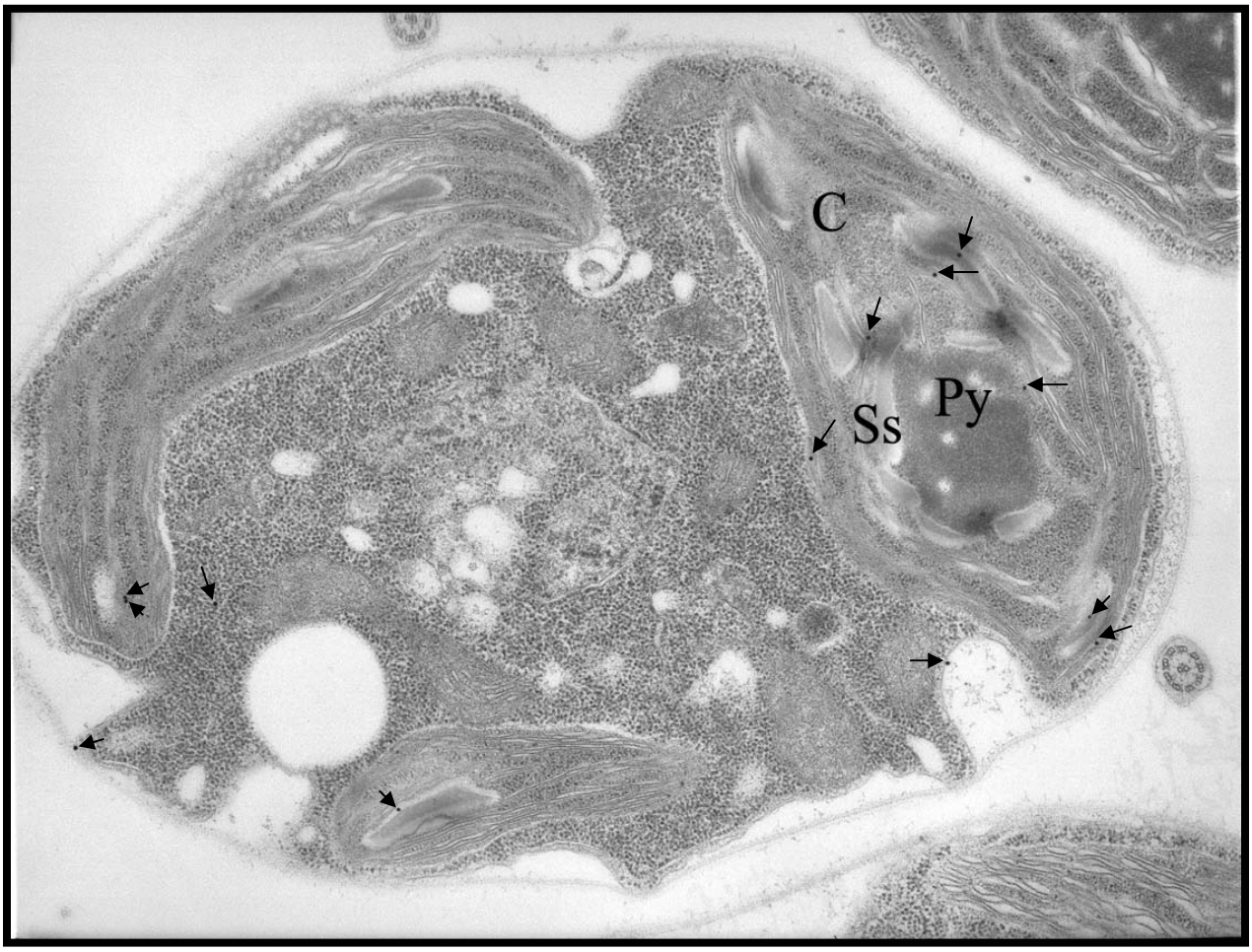



**Figure 4.17 A transmission electron micrograph showing the immunogold labeling of *C. reinhardtii* cells probed with the Cah6 antibody.** Cells grown under low $CO_2$ (0.03%) conditions in minimal medium were probed with the Cah6 antibody. Ss, Py and C denote starch sheath, pyrenoid and chloroplast, respectively. Immunogold labelings are shown by black arrows.

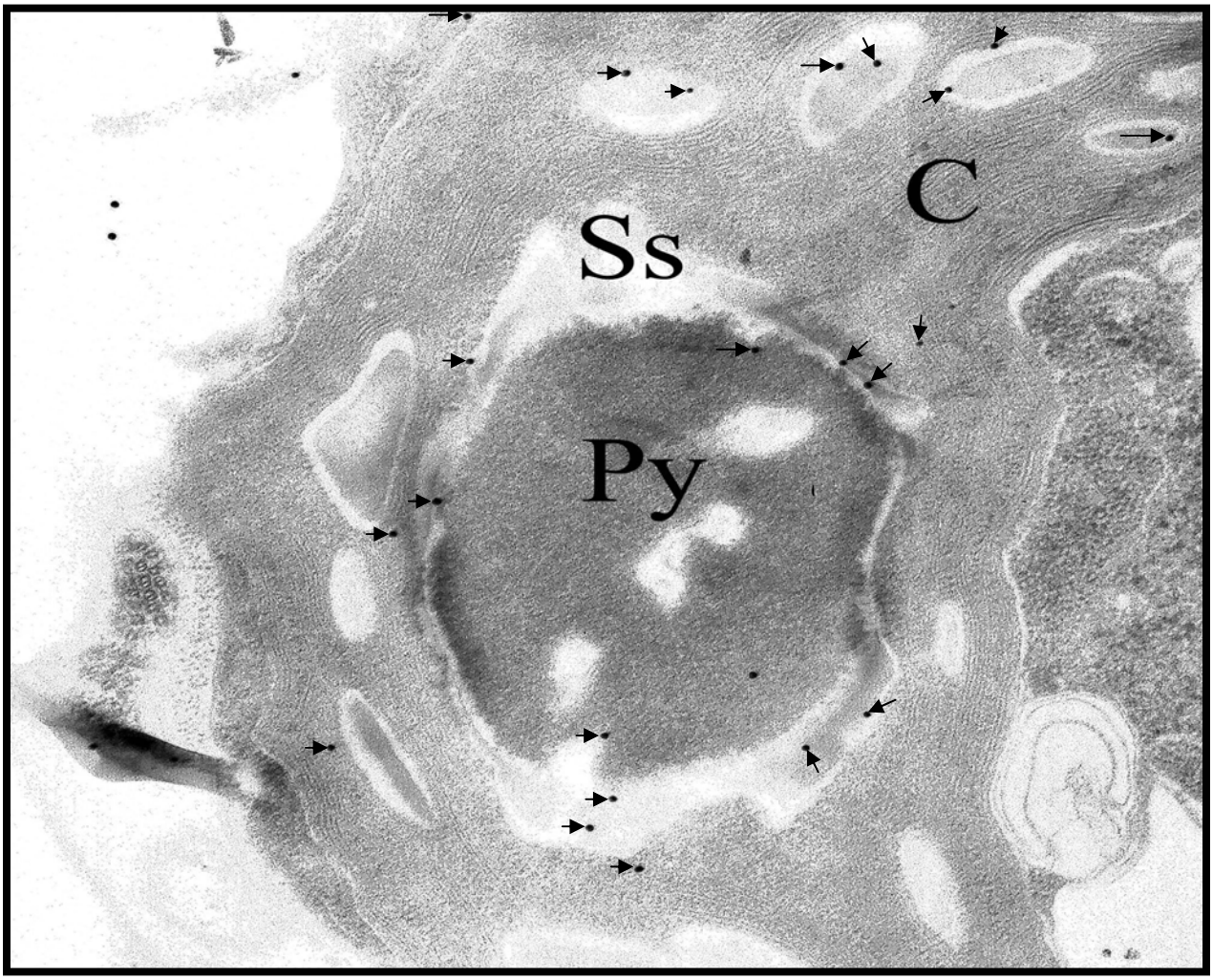



**Figure 4.18 A transmission electron micrograph showing the immunogold density around the pyrenoid in *C. reinhardtii* cells probed with the Cah6 antibody.** Cells grown under low $CO_2$ (0.03%) conditions in minimal medium were probed with the Cah6 antibody. Ss, Py and C denote starch sheath, pyrenoid and chloroplast, respectively. Immunogold labelings are shown by black arrows.


**Figure 4.19 A transmission electron micrograph showing the immunogold labeling of *C. reinhardtii* cells probed with the preimmune serum.** Low $CO_2$ (0.03%) adapted cells grown in minimal medium were used. SS, P and C denote starch sheath, pyrenoid and chloroplast, respectively. Immunogold labelings are shown by small black arrows.

**Table 4.2** The intracellular localization of the Cah6 using wild type *C. reinhardtii* cells

| Location | Area ($\mu m^2$) | Immunogold density (# of immunogold particles / $\mu m^2$) | | |
|---|---|---|---|---|
| | | Immune | Preimmune | Difference |
| Outside | 7.80 ± 0.38 | 1.67 ± 0.10 | 1.54 ± 0.11 | 0.13 ± 0.19 |
| Cytoplasm | 6.84 ± 1.2 | 0.44 ± 0.02 | 0.30 ± 0.02 | 0.14 ± 0.03 |
| Nucleus | 0.36 ± 0.05 | 0.47 ± 0.02 | 0.47 ± 0.02 | 0.00 ± 0.01 |
| Stroma | 5.93 ± 1.07 | 3.0 ± 0.14 | 0.85 ± 0.04 | 2.14 ± 0.13 |
| Pyrenoid | 1.85 ± 0.30 | 0.99 ± 0.08 | 0.97 ± 0.7 | 0.02 ± 0.03 |
| Starch sheath | 0.86 ± 0.11 | 10.9 ± 1.9 | 2.3 ± 0.14 | 8.6 ± 1.95 |

Immunogold densities in different cell compartments in sections were calculated by dividing the number of immunogold particles in a particular cell organelle with the area of that cell organelle. The data presented in the table is the average ± SD of 15 cell sections. The stroma includes the thylakoid area. The cytoplasmic area was calculated by subtracting the total area of the chloroplast and nucleus from the cell area. The Cah6 antibody dilution used for immunolocalization was 1:10. The cells used for immunolocalization are air adapted D66 cells grown in minimal medium.

**Table 4.3** The inhibition constants of the bovine CAII and heterologously produced Cah6

| Inhibitor | $I_{50}$ of bovine CA (M) | $I_{50}$ of Cah6 (M) |
|---|---|---|
| Acetazolamide | $1.4 \times 10^{-8}$ | $2 \times 10^{-6}$ |
| Ethoxyzolamide | $1.2 \times 10^{-9}$ | $9 \times 10^{-6}$ |
| Azide | $1.1 \times 10^{-3}$ | $1.5 \times 10^{-5}$ |
| Cyanide | $4.9 \times 10^{-5}$ | $5 \times 10^{-6}$ |

Erythrocyte bovine CAII was purchased from Sigma. The $I_{50}$ value corresponds to the concentration giving 50% inhibition. $I_{50}$ was determined by plotting percentage of inhibition vs. concentration of inhibitor. Sodium salts of azide and potassium salts of cyanide were used. The data shown for each treatment in the table are the averages ± SD of three different measurements.

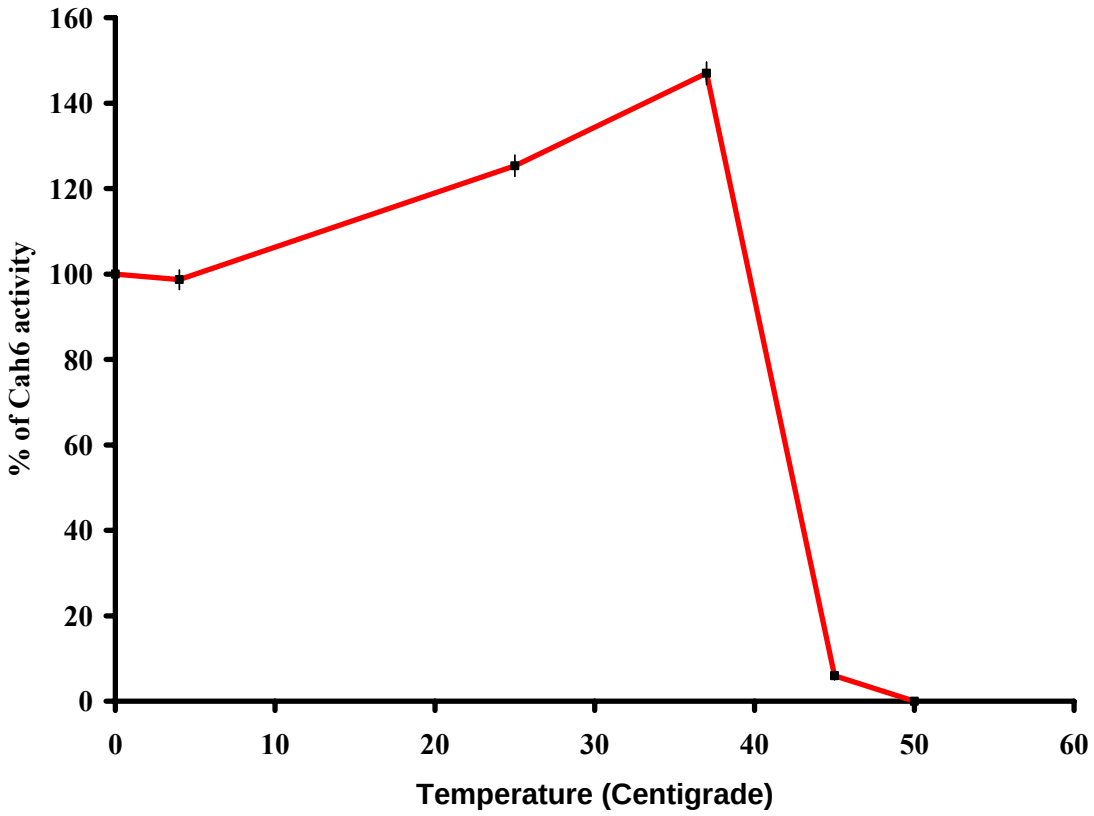



**Figure 4.20 The thermostability of Cah6 activity.** Cah6 was incubated for 15 minutes at the indicated temperatures and cooled on ice. CA activity was measured at 4°C by the Wilbur-Anderson method. Activity is represented as a percentage of the activity of a sample maintained on ice throughout the experiment (400 WAU/mg). The percentage of Cah6 activity data for each temperature treatment shown in the graph are the averages ± SD of three sample evaluations.

**Table 4.4** The effects of the SH-reducing agents on Cah6 activity

| SH- reducing agents | % of Cah6 activity |
|---|---|
| Ox-CA | 100 ± 0 |
| 2-Mercaptoethanol | 100 ± 0 |
| Cysteine | 100 ± 0 |
| Dithiothreitol | 100 ± 0.58 |

Cah6 was purified without any reducing agents in the buffer and the activity was measured. The specific activity (400U/mg) in the oxidized state is denoted as 100% activity (ox-CA). Prior to the measurements of CA activity, purified recombinant Cah6 enzyme was incubated for 30 minutes at room temperature with: 10 mm 2-mercaptoethanol, 10 mM Cysteine and 10 mM dithiothreitol. The data shown for each treatment in the table are the averages ± SD of three different measurements.

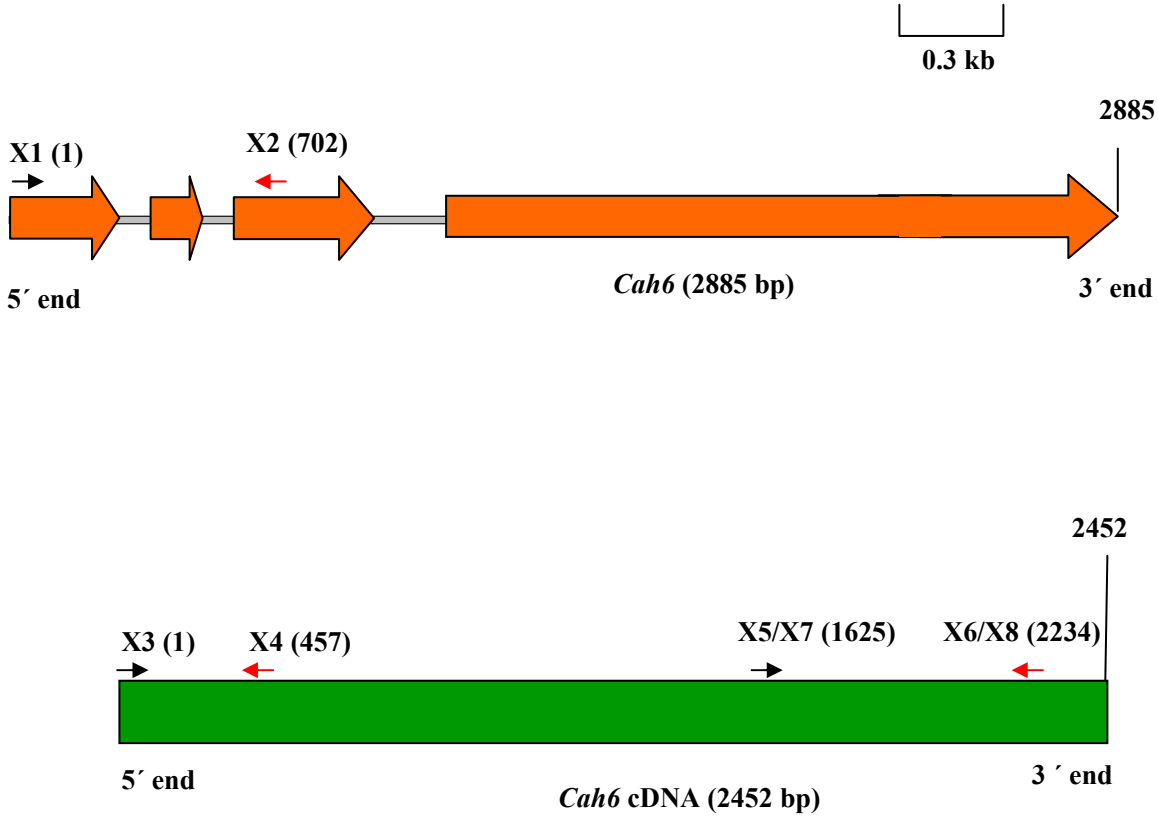



**Figure 4.21 The genomic and cDNA maps of *Cah6* showing the alignment of *Cah6* RNAi primers.** The orange block arrows and the lines interconnecting them represent the four exons and three introns, respectively. 5′ end and 3′ end RNAi primers are shown by black and red arrows, respectively. The numbers within the parentheses beside the primers denote the positions of the primers in base pairs.

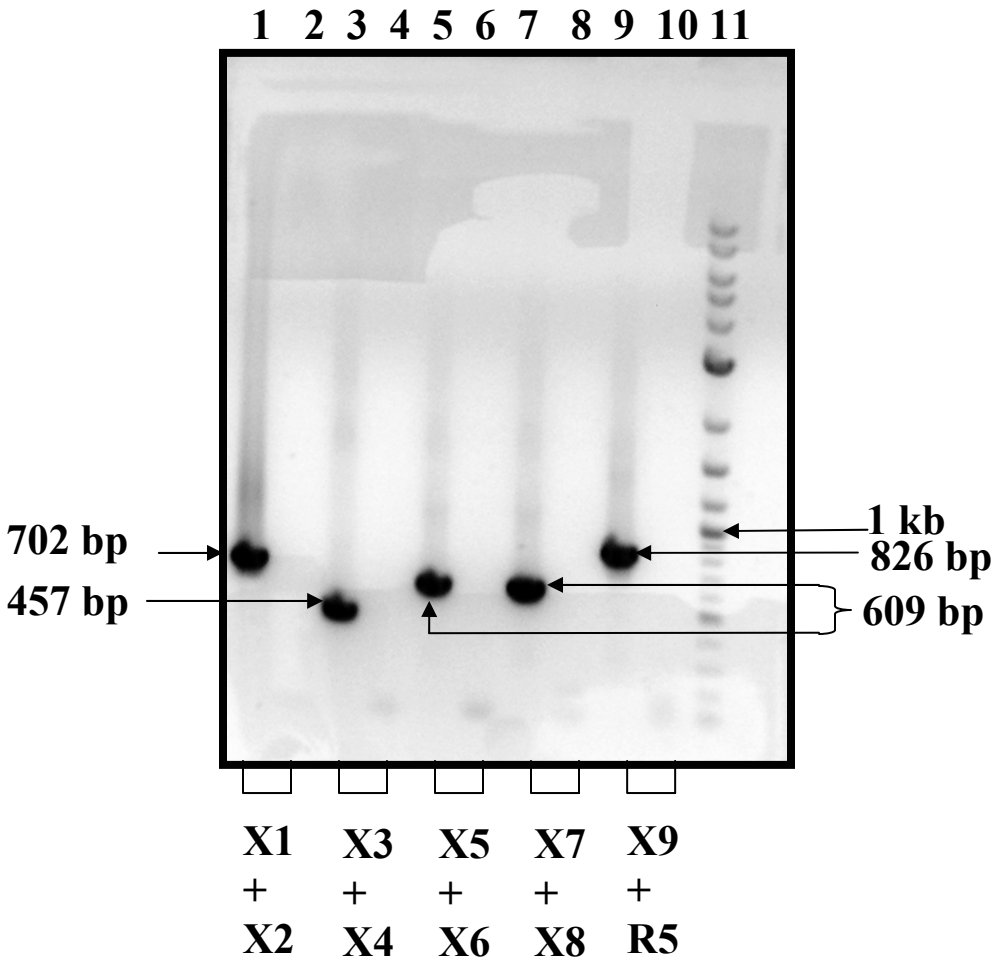



**Figure 4.22 PCR product generated using the *Cah6* RNAi primers.** The PCR primers are given at the bottom of each lane. Lane 1 is the PCR product using the cosmid 72-E-6 DNA. Lanes 2, 4, 6, 8 and 10 are the PCR results in the absence of DNA. Lanes 3, 5, 7, 9 represent PCR using the cDNA core library. X9 and R5 are the primers that were used to amplify an 826 bp cDNA product. This product was used as a probe for *Cah6* Northern blot analyses. Lane 11 represents 1 kb ladder (NEB).

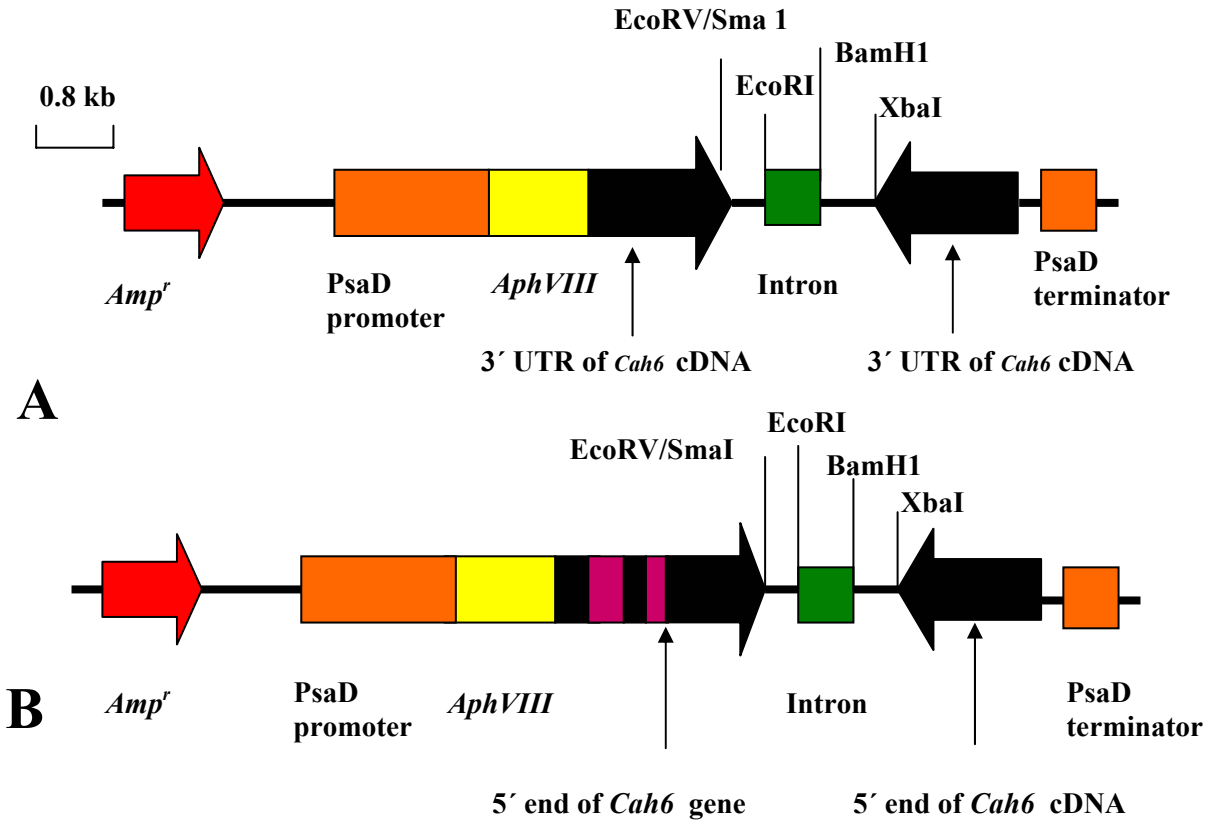



**Figure 4.23 RNAi constructs of *Cah6*.** A. 3′end RNAi construct. B. 5′end RNAi construct. The reverse complementary fragments of Cah6 were cloned in the pSL72 vector which is 5495 bp long. Restriction enzymes used to cut the vector and the Cah6 DNA fragments are shown by black lines. *AphVIII* and *Amp*$^r$ codes for paromomycin and ampicillin resistance, respectively. PsaD is a constitutive promoter. The intron represented by the green box is the second intron of the cytochrome $c_6$ gene of *C. reinhardtii*. Pink boxes in the 5′ end of Cah6 gene represent two introns.

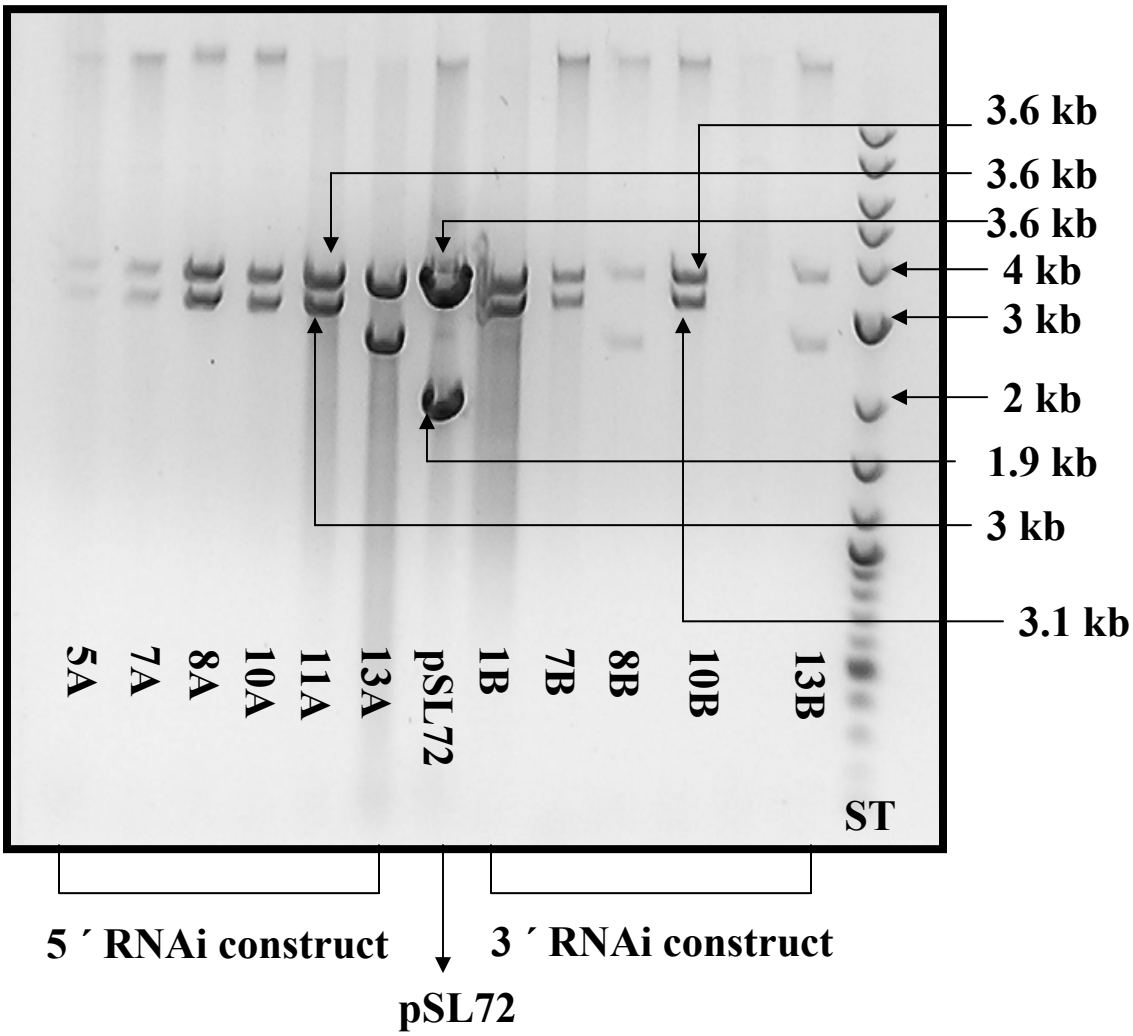



**Figure 4.24 The restriction enzyme digestions of *Cah6* RNAi clones.** AvrII and XbaI were used for double restriction digestion. The RNAi clone number is denoted by a number followed by an alphabet. "A" and "B" represent 5´ and 3 ´ RNAi constructs respectively. ST represents the DNA 2-log ladder (NEB).

rate of photoautotrophic growth (Giordano *et al*., 2003). The expression of these CAs was induced at 0.2% $CO_2$ condition by increasing the $NH_4^+$ concentration in the growth medium. These workers have proposed that the mitochondrial CAs are involved in supplying $HCO_3^-$ for anaplerotic assimilation catalyzed by phosphoenolpyruvate carboxylase, which in turn provides carbon skeletons for nitrogen assimilation under certain conditions.

Other examples include the fact that deletion of the β-CA-like gene *NCE103* of the yeast *S. cerevisiae* causes an oxygen-sensitive growth defect (Gőtz *et al*., 1999). It has also been found that the tobacco salicylic acid-binding protein 3 (SABP3) is a chloroplast β-CA which exhibits antioxidant activity and plays a role in the hypersensitive defense response (Slaymaker *et al*., 2001). In *E. coli*, the *cynT* gene which codes for a β-CA is a part of the *cyn* operon; this β-CA recycles the $CO_2$ produced in the reaction of cyanate with $HCO_3^-$, back to $HCO_3^-$, which would have diffused out of the cell otherwise (Guilloton *et al*., 1993). Finally, a β-CA gene identified in *Corynebacterium glutamicum* has been shown to be essential for achieving normal growth under atmospheric conditions (Mitsuhashi *et al*., 2003). These workers have shown that the effect of this β-CA is most likely due to its ability to maintain favorable intracellular $HCO_3^-$ levels, particularly during exponential growth phases and also in the case of L-lysine overproduction, both of which are conditions of higher $HCO_3^-$ demand (Mitsuhashi *et al*., 2003).

*Cah6* is the third β-CA and the sixth CA gene to be identified in *C. reinhardtii*. The Cah6 protein has two cysteine and one histidine residues as zinc coordinating ligands, similar to all known enzymatically active β-CAs (Moroney *et al*., 2001). It is less susceptible to sulfonamide inhibitors than the α-CA from bovine erythrocytes, a characteristic trait of all active β-CAs. In general, β-CAs have sulfonamide $I_{50}$ ranging from 2 μM to10 μM. The $I_{50}$ of Cah6 falls within this range. The full length protein has a calculated molecular weight of 28 kDa and a pI of 7.0 (predicted by various protein prediction programs listed under the ExPASY website). It

has an apparent weight of 31 kDa on a SDS-polyacrylamide gel. It is similar to β-CAs from *E. coli*, green algae and higher plants with an amino acid identity of 23% to 34%, and most closely resembles the green alga *Coccomyxa* β-CA with a 34% identity. Like the activity of the *Coccomyxa* β-CA, that of Cah6 is unaffected by SH-reducing agents. The enzyme lost more than 50% of its activity above 43°C and became inactive at 50°C, with an optimal activity around 33°C. Thus the optimal temperature for Cah6 activity, like that of Cah3, falls within the range of the optimal growth temperature (12°C-35°C) of *C. reinhardtii*. The mature Cah6 protein (lacking the 39 amino acid residues that comprise the signal peptide of Cah6) has a calculated molecular weight of 24 kDa and a pI of 6.58. It has an apparent molecular weight of 28 kDa on a SDS-polyacrylamide gel.

CA activity was detected in cell extracts of the induced B48 clone but not in that of the B3 clone of *E. coli*, expressing the Cah6 gene. The B48 clone contains the entire ORF of Cah6 whereas the B3 clone had only the cDNA coding for the mature Cah6 protein. Eukaryotic proteins overexpressed in *E. coli* cells usually do not undergo post translational modifications, seen in a eukaryotic cell. The mature Cah6 protein (without the transit peptide) expressed in the B3 clone had a higher chance of getting properly folded than that (full length protein) expressed in the B48 clone. Under this circumstance protein from the B3 clone would be expected to be enzymatically more active than that from the B48 clone. If the overexpression of the Cah6 is toxic to *E. coli* cells, bacteria can either degrade the overexpressed protein or specifically introduce mutations in the DNA sequence coding for the toxic protein (personal communication, Dr. S.G. Bartlett, Louisiana State University, Baton Rouge, Louisiana). Thus the original B3 clone that I had selected by DNA sequencing (before measuring activity assays), could have got lost in the course of selection of the clones that had the mutated Cah6 protein. This can explain the lack of CA activity in the protein from B3 clone. In an effect to confirm this hypothesis,

resequencing of the *Cah6* gene in the B3 clone will be necessary in order to eliminate the possibility for any mutations introduced by the bacteria.

Alternatively, there can be another explanation for the lack of activity of the protein from B3 clone. The first amino acid in the mature Cah6 is an arginine. Cloning of the Cah6 cDNA coding for mature Cah6 in B3 clone was done by eliminating the codon for arginine as Factor Xa does not cleave any protein that starts with an arginine after the Factor Xa recognition sequence [ Ile-(Glu-Asp)-Gly-Arg]. Thus mature Cah6 protein in the B3 clone started with a serine instead of an arginine residue. If no mutations are found in the *Cah6* sequence in B3 clone, then site directed mutagenesis experiments may be used to alter the arginine residue in the enzymatically active Cah6 from the B48 clone in order to study whether or not the single arginine residue in Cah6 does play a critical role in modulating the catalytic activity of Cah6.

Western and Northern blotting analyses show that Cah6 is constitutive and slightly upregulated under low $CO_2$ conditions. Immunolocalization shows that Cah6 is localized in the stroma of the chloroplast and is not present in the pyrenoid. Interestingly, in *C. reinhardtii* immunogold density is four fold higher in the area around the pyrenoid, particularly around the starch sheath which surrounds the pyrenoid, compared to that in the other areas in the stroma of the chloroplast (Table 4.2). Based on these observations, I hypothesize that Cah6 indirectly plays a role in the CCM by trapping $CO_2$ diffusing out from the pyrenoid, the site of localization of Rubisco in *C. reinhardtii*, and converting it to $HCO_3^-$ (Fig. 4.25). This conversion of $CO_2$ to $HCO_3^-$ would increase the $HCO_3^-$ pool in the stroma, thereby promoting the Rubisco activity inside the pyrenoid. Generation of Cah6 mutants might provide insight into the role of Cah6 in the chloroplast of *C. reinhardtii*. At this point, none of the six RNAi air acclimated transformants screened show any visible difference in the expression of Cah6 protein in the cells compared to that in the wild type cells. Hence no deduction can be made about the functional

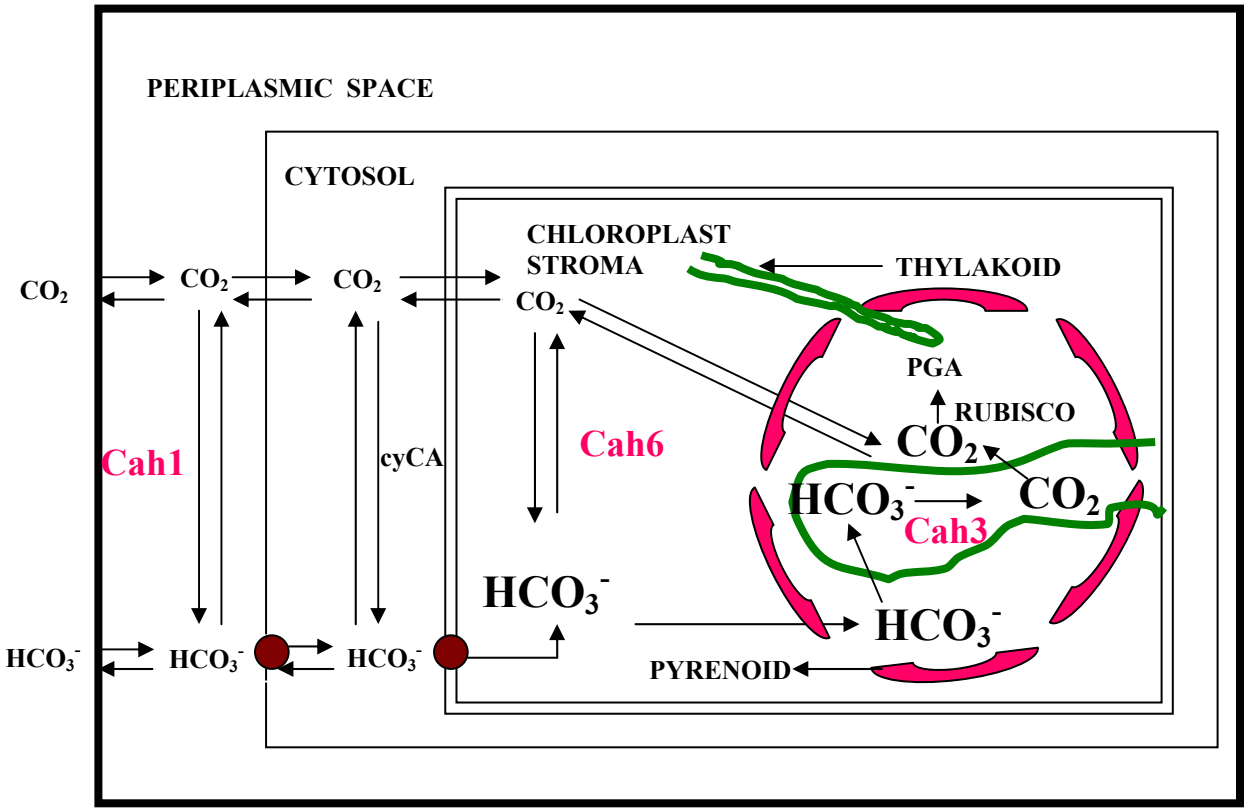



**Figure 4.25 A model showing the potential role of Cah6 and other known CAs in the operation of CCM in *C. reinhardtii*.** The font sizes of $CO_2$ and $HCO_3^-$ indicate the relative concentrations of these Ci species. cyCA represents cytoplasmic carbonic anhydrase. Cah1 and Cah3 represent the periplasmic and thylakoid CAs, respectively. Putative $HCO_3^-$ transporters are denoted by small brown circles.

role played by Cah6 in the CCM and photosynthesis. Further screening of the remaining air adapted transformants by Western blotting will provide definite indication if the RNA interference is in operation in the wild type *C. reinhardtii* cells.

# CHAPTER 5

# IDENTIFICATION, CLONING AND OVEREXPRESSION OF A γ-CA LIKE GENE (*Gclp 1*)

## INTRODUCTION

The first γ-CA discovered was from the archaebacterium *Methanosarcina thermophila* (Alber and Ferry, 1994). Since that time, although genes encoding putative γ-CA proteins have been found in eubacteria and plants (Newman *et al*., 1994), γ-CA from *M. thermophila* is the only enzymatically active γ-CA known to date. One of the carboxysomal shell proteins in *Synechococcus* PCC7942 (Price *et al*., 1993), CcmM, has been found to have a strong N-terminal sequence homology with the CA active site from γ-CAs. The C-terminal region of CcmM has three to four 87 amino acid repeats that are very similar to the small subunit of Rubisco protein from the cyanobacteria. Cells deleted in *CcmM* show a HCR phenotype and have empty carboxysomes. From these results it is clear that CcmM is required for correct carboxysome assembly and plays an essential role in the CCM in cyanobacteria. It is not clear, however whether CcmM has CA activity, or if its enzymatic activity is needed for correct assembly of carboxysomes. Three *Arabidopsis* ESTs in the databases have homologies with the γ-CA from *M. thermophila*, but it is not yet known if any of the *Arabidopsis* γ-CA proteins have CA activity or what their physiological roles might be. More information about γ-CAs may be found in Chapter 1.

The *Chlamydomonas* EST database has two contigs that align well with the γ-CA of *M. thermophila* (Cam) and the cyanobacterial γ-CA homologue CcmM. These two genes encoding γ-CA like proteins have been termed *Gclp1* and *Gclp2*, respectively. *Gclp1* and *Gclp2* have more than 50 and 40 ESTs, respectively, implying that they are highly expressed. In this chapter

I report my research on the *Gclp1* gene. This is the first report of a putative γ-CA gene in *C. reinhardtii*.

**RESULTS**

**1. Screening of Cosmid and cDNA Libraries for Genomic and cDNA Clones of *Gclp1***

The γ-CA protein sequence of *M. thermophila* was used to BLAST the EST database of *C. reinhardtii*. The search yielded several ESTs that were analyzed by CAP (contig assembly program; http://www.infobiogen.fr/services/analyseq/cgi-bin/cap_in.pl) to form the consensus *Gclp*1 and *Gclp2* sequences. Gclp1 was selected for further studies. In order to amplify *Gclp1*, several PCR primers were designed based on the *Gclp1* contig. The PCR primers, 1F and 2B, were used with genomic DNA to generate a product of 745 bp. This PCR product was used to make a probe to screen an indexed cosmid library. After three sequential rounds of screening, two cosmid clones, named 70-C-3 and 70-G-8, were isolated. Positive cosmid clones were verified as *Gclp1* clones, after each round of screening, by PCR followed by sequencing of the PCR product using different *Gclp1* primers. Primers M1F and M1R were used on the two cosmid clones to yield a PCR product of 2249 bp (Fig. 5.1). The same two primers used on a cDNA core library yielded a PCR product of 1170 bp (Fig. 5.1).

Detailed information on the screening of the cosmid library and preparation of cDNA library may be found in Chapter 2 of this dissertation. Sequences and the alignment of all *Gclp1* primers are given in Appendix 3 of this dissertation.

**2. Sequencing and Homology Search**

The cosmid clones 70-C-3 and 70-G-8, and the 1170 bp cDNA PCR product mentioned above, were sequenced in both directions using M1F and M1R primers. The sequencing results matched those in the *Chlamydomonas* EST and genomic databases. The *Gclp1* gene has seven exons and six introns and is 2374 bp long (Fig. 5.2). The size ranges of exons and introns are

from 69 bp to 583 bp and 103 bp to 429 bp, respectively. The full genomic sequence is given in Fig. 5.3. The *Gclp1* cDNA is 1294 bp long (Fig. 5.4). A restriction map of the full length cDNA clone of *Gclp*1 is shown in Fig. 5.5. *Gclp1* cDNA encodes a protein of 312 amino acids. It contains a translation start site at nucleotide 104 and a stop site at nucleotide 1039 (Fig. 5.6). Predictions based on various protein prediction programs (SORT P, CHLOR P AND TARGET P) listed under ExPasy server (http://ca.expasy.org/tools/#translate), indicate that it is either located in the cytoplasm or is secreted outside the cell. In the latter case, a putative transit peptide of 41 amino acids has been predicted. Gclp1 has 30%-40% homologies with γ-CA of *M. thermophila* and CcmM of cyanobacteria, and contains the characteristic three histidines which may serve as zinc coordination residues seen in the enzymatically active γ-CA of *M. thermophila* (Fig. 5.7).

**3. Cloning of *Gclp1* in an Overexpression Vector**

*Gclp1* was cloned in the overexpression vector pMal-c2x to study Gclp1 properties and to raise an antibody, if the Gclp1 was found enzymatically active. Two different pMal-*Gclp1* recombinant overexpression vectors were constructed. One contained the cDNA sequence coding for the full length open reading frame (ORF) of the Gclp1 protein and the other contained the cDNA sequence that would code for the mature Gclp1 protein, if the first 41 amino acids were a leader sequence as predicted by the various protein prediction programs.

S1 + M1RH primers and S3 + MIRH primers were used to amplify the cDNA coding for the ORF and putative mature Gclp1 protein, respectively (Fig. 5.8). Amplified cDNAs were purified from the gel and cloned into the pMal vector (Fig. 5.9) to generate two different types of recombinant *MalE-Gclp1* constructs as described in Chapter 2. The in-frame insertion of *Gclp1* with the sequence of MBP in the recombinant clone was verified by restriction enzyme digestion analyses (Fig. 5.10) followed by DNA sequencing of the *Gclp1* cDNA sequence cloned

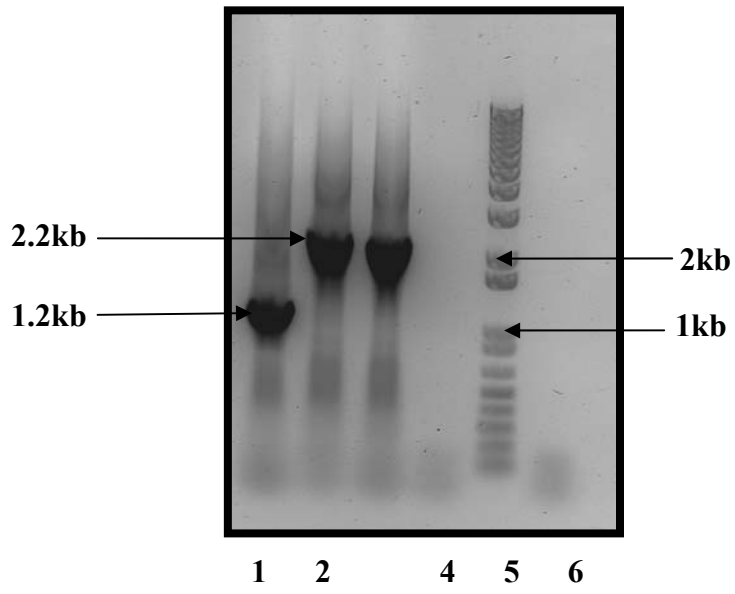



**Figure 5.1 The screening of the cDNA and cosmid libraries for a cDNA and genomic *Gclp1* clone using PCR.** Lane 1 shows the PCR result using M1F and M1R γ-CA primers on the cDNA library. Lane 2 and Lane 3 show the PCR results using MIF and MIR primers on the genomic cosmids 70-C-3 and 70-G-8, respectively. Lane 4 and 6 represent PCR products in the absence of DNA. Lane 5 contains a 1kb DNA ladder.

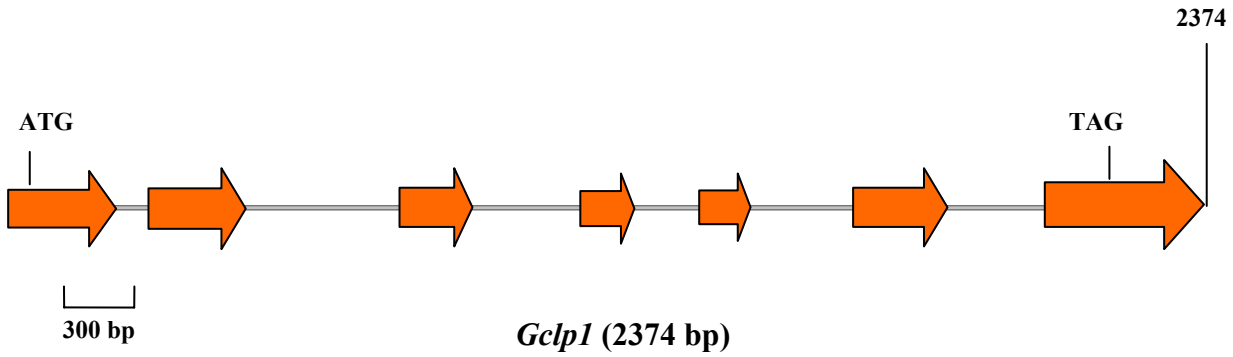



**Figure 5.2 The genomic map of *Gclp1*.** The orange block arrows and lines connecting the block arrows represent the seven exons and six introns, respectively. The start and stop codons are labeled by black lines. The numbers within the parenthesis denote the nucleotide positions of the start and stop codons. *Gclp1* is 2374 base pairs long.

GCAGTTAGTCCATGCGAATCGTCTGGCCTTGTGCGGGGCTTTCTACGTGACACAAGGATTTAGG
CCCTGAGTGCCCGATAGCTTTGCCCCAAAGCGGTGGACC**ATG**TCGCTATTCAAGTCTAGCCTGC
CTGCGGGCTTCCTATTCCCCTATCGGCACCCCAAGGCCAAGGGGCTTGTTGAGGGCACGCTTTA
TGGACTGGGCTCCCTGTTTCGCGGCGTGGGCGCCGCGCTGGATGAGCTGGGCTCTATGGTTCAG
GGCCCTCAGGGTAGTGTCAAGGACCACGGTGCGTTTGCAGGACAGTGGGCGTGCTGGAACACGT
TAATCGCATCCTTGCTTCGTTCTGCTTACTGCGCGCAGTCCAGCCTAACCTGGCGTTTGCACCA
GTGCACCGCAAGCCGGATGTGCCCGTTAACGCGGGCCAGGTGGTGCCCGCTCCACCCGCTGCTG
CTCGCACGCTGAAAATCAAGGAGGTGGTTGTGCCCAACAAGCACAGCACCGCGTTCGTGGCTGC
CAACGCCAATGTGCTCGGGAACGTTAAGCTGGGGGCGGGCTCATCGGTGTGGTATGGCGCCGTG
CTGCGCGgtgagtgtgtgatacgcagaggggggaggggaagaaaggcgaggtagcaggtcgggt
ggtcggggtcgcgggctttggttttcgacatgggtgggcggcgtcctgtgtttgtggccatgac
gaacgggcatacgcgtagtcgttgtgggggcgtgtgggggtgcgctttgaacggtctatctgcg
ctcgccggcctacccaatgccgttgcgtatgcggcgaccgggggagtctaaacccatgcgaaat
acaccagggtgtattggtcccaaacttgtggattgctgcagcaagggaagtggctgtatcattg
ggacgaaatgcaatgagtgggctcaagcgtcaggaggacgttttggcaggcgacgcctcgcgac
gcagtaccacatatgggcaggccgctgacgcgtacccgcttccgcccaccagGTGACGTGAACG
GCATTGAGGTGGGCGCCAACAGCAACATCCAGGACAACGCCATCGTGCACGTGTCCAAGTACAG
CATGGACGGCACGGCACGGCCCACCGTCATCGGCAACAATGTGACCATTGgtgaggcacgccgt
gagggatccgggggggggggggggaaangcacaaacgcgtcttcatactggcatgaaagggttc
cgcgtctacgctggcgcgcaaacatcacggtacgcgacgcgacgctcacgtagattctgttctt
tgggtctttgcagtgccacggcgccacgcgtgttagaacctttgtgtctcacacaccccatccc
caaccccacacactcagGCCACGCCGCCACGGTGCACGCCTGCACCATTGAGGACAACTGCCTG
GTGGGCATGGGCGCCACCGTGCTCGACGGAGCGACGgtgagggagcggggcggaagccatggtg
ctggtcgcttgtgtatgtgccatctatctcttgaggtctgataaaccgcatgtacctccgctta
cgcacggacagGTCAAGAGCGGCTCCATCGTGGCTGCCGGCGCCGTGGTGCCGCCCAACACCAC
CATCCCCTCGGGCCAGgtgagccgcgctctgcgcccagctttgccgtacatctagtgcgtggtg
ctctgcatctaaatgcaagctgcggttgacgcaagcaattgcttgttatgggcggccattgaac
gtggtgcatacggattgtgccttggtccagGTGTGGGCCGGCTCGCCCGCCAAGTTCCTGCGCC
ACCTGGAGCCGGAGGAGGCCAGCTTCATCGGCAAGTCTGCCAGCTGCTACGCCGAGCTGTCCGC
CATCCACAAGTgtgagtgcggctcgctgtcgtcctctcgcgcacgtcctgtgtgtctggttgtc
tttgccgtctagcatccgccatgccaacccttgcaccttcccccaacgtcctgcagTCGAGCAG
AGCAAGACGTTTGAGGAGCAGTACACGGAGAGCTGCATCATCAAGGACCGCGCCGCTCTGGCCG
ACCCGTCAAACTCAGTGCACCAGATGTGGGAGTACGACAGCCAGACGGCGTTGGTGGCCCGCGC
CAAGAGG**TAG**AGAGAGGAGCTGCAAGGGCTGCTGCGGTAGGCATGCACTGCAGGCAGGCAGGGT
TGCCATGGCAGCAGTCGAGGTGAACGACACGCGCGTAGGCTGTCATGCGTTCGGCGTTGAATGA
GTACGGTGCACCGCACCAACAGCGGCTAATTAAAGTAGCGGAAGTATAGTTGAGACAACCACAG
ATTTAAGTAATGTAGCGTCAATGTACCGTGGGGCATGGGCGCCCCAAACCTATTTAAATGACTC
CCTATC

**Figure 5.3 The genomic sequence of *Gclp1*.** The Gclp1 gene is 2374 bp long. The exon sequences are in upper case and the intron sequences are in lower case. The start and stop codons are shown in bold red.

GCAGTTAGTCCATGCGAATCGTCTGGCCTTGTGCGGGGCTTTCTACGTGACACAAGGATTTAGG
CCCTGAGTGCCCGATAGCTTTGCCCCAAAGCGGTGGACC**ATG**TCGCTATTCAAGTCTAGCCTGC
CTGCGGGCTTCCTATTCCCCTATCGGCACCCCAAGGCCAAGGGGCTTGTTGAGGGCACGCTTTA
TGGACTGGGCTCCCTGTTTCGCGGCGTGGGCGCCGCGCTGGATGAGCTGGGCTCTATGGTTCAG
GGCCCTCAGGGTAGTGTCAAGGACCACGTCCAGCCTAACCTGGCGTTTGCACCAGTGCACCGCA
AGCCGGATGTGCCCGTTAACGCGGGCCAGGTGGTGCCCGCTCCACCCGCTGCTGCTCGCACGCT
GAAAATCAAGGAGGTGGTTGTGCCCAACAAGCACAGCACCGCGTTCGTGGCTGCCAACGCCAAT
GTGCTCGGGAACGTTAAGCTGGGGGCGGGCTCATCGGTGTGGTATGGCGCCGTGCTGCGCGGTG
ACGTGAACGGCATTGAGGTGGGCGCCAACAGCAACATCCAGGACAACGCCATCGTGCACGTGTC
CAAGTACAGCATGGACGGCACGGCACGGCCCACCGTCATCGGCAACAATGTGACCATTGGCCAC
GCCGCCACGGTGCACGCCTGCACCATTGAGGACAACTGCCTGGTGGGCATGGGCGCCACCGTGC
TCGACGGAGCGACGGTCAAGAGCGGCTCCATCGTGGCTGCCGGCGCCGTGGTGCCGCCCAACAC
CACCATCCCCTCGGGCCAGGTGTGGGCCGGCTCGCCCGCCAAGTTCCTGCGCCACCTGGAGCCG
GAGGAGGCCAGCTTCATCGGCAAGTCTGCCAGCTGCTACGCCGAGCTGTCCGCCATCCACAAGT
TCGAGCAGAGCAAGACGTTTGAGGAGCAGTACACGGAGAGCTGCATCATCAAGGACCGCGCCGC
TCTGGCCGACCCGTCAAACTCAGTGCACCAGATGTGGGAGTACGACAGCCAGACGGCGTTGGTG
GCCCGCGCCAAGAGG**TAG**AGAGAGGAGCTGCAAGGGCTGCTGCGGTAGGCATGCACTGCAGGCA
GGCAGGGTTGCCATGGCAGCAGTCGAGGTGAACGACACGCGCGTAGGCTGTCATGCGTTCGGCG
TTGAATGAGTACGGTGCACCGCACCAACAGCGGCTAATTAAAGTAGCGGAAGTATAGTTGAGAC
AACCCCAGATTTAAGTAA**TGTAG**CGTCAATGTACCGTGGGGCATGGGCGCCCCAAACCTA**TTTA**
**A**ATGACTCCCTATCAAAAAAAAAAAA

**Fig 5.4 The full length cDNA sequence of *Gclp1*.** The length of the cDNA is 1294 base pairs (excluding the Poly A tail). The start and stop codons are shown in bold red. The putative polyadenylation signals are in bold blue.

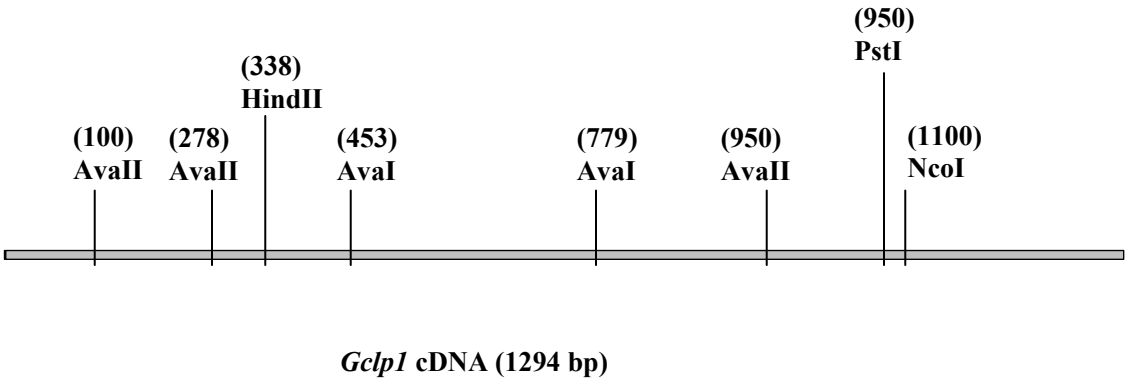



**Figure 5.5 The restriction map of the *Gclp1* cDNA.** The restriction enzyme sites are indicated by black lines. The nucleotide numbers within the parentheses indicate positions of the cleavage sites of enzymes in base pairs.

```
109 atgtcgctattcaagtctagcctgcctgcgggcttcctattcccc
    M  S  L  F  K  S  S  L  P  A  G  F  L  F  P
154 tatcggcaccccaaggccaaggggcttgttgagggcacgctttat
    Y  R  H  P  K  A  K  G  L  V  E  G  T  L  Y
199 ggactgggctccctgtttcgcggcgtgggcgccgcgctggatgag
    G  L  G  S  L  F  R  G  V  G  A  A  L  D  E
244 ctgggctctatggttcagggccctcagggtagtgtcaaggaccac
    L  G  S  M  V  Q  G  P  Q  G  S  V  K  D  H
289 gtccagcctaacctggcgtttgcaccagtgcaccgcaagccggat
    V  Q  P  N  L  A  F  A  P  V  H  R  K  P  D
334 gtgcccgttaacgcgggccaggtggtgcccgctccacccgctgct
    V  P  V  N  A  G  Q  V  V  P  A  P  P  A  A
379 gctcgcacgctgaaaatcaaggaggtggttgtgcccaacaagcac
    A  R  T  L  K  I  K  E  V  V  V  P  N  K  H
424 agcaccgcgttcgtggctgccaacgccaatgtgctcgggaacgtt
    S  T  A  F  V  A  A  N  A  N  V  L  G  N  V
469 aagctgggggcgggctcatcggtgtggtatggcgccgtgctgcgc
    K  L  G  A  G  S  S  V  W  Y  G  A  V  L  R
514 ggtgacgtgaacggcattgaggtgggcgccaacagcaacatccag
    G  D  V  N  G  I  E  V  G  A  N  S  N  I  Q
559 gacaacgccatcgtgcacgtgtccaagtacagcatggacggcacg
    D  N  A  I  V  H  V  S  K  Y  S  M  D  G  T
604 gcacggcccaccgtcatcggcaacaatgtgaccattggccacgcc
    A  R  P  T  V  I  G  N  N  V  T  I  G  H  A
649 gccacggtgcacgcctgcaccattgaggacaactgcctggtgggc
    A  T  V  H  A  C  T  I  E  D  N  C  L  V  G
694 atgggcgccaccgtgctcgacggagcgacggtcaagagcggctcc
    M  G  A  T  V  L  D  G  A  T  V  K  S  G  S
739 atcgtggctgccggcgccgtggtgccgcccaacaccaccatcccc
    I  V  A  A  G  A  V  V  P  P  N  T  T  I  P
784 tcgggccaggtgtgggccggctcgcccgccaagttcctgcgccac
    S  G  Q  V  W  A  G  S  P  A  K  F  L  R  H
829 ctggagccggaggaggccagcttcatcggcaagtctgccagctgc
    L  E  P  E  E  A  S  F  I  G  K  S  A  S  C
874 tacgccgagctgtccgccatccacaagttcgagcagagcaagacg
    Y  A  E  L  S  A  I  H  K  F  E  Q  S  K  T
919 tttgaggagcagtacacggagagctgcatcatcaaggaccgcgcc
    F  E  E  Q  Y  T  E  S  C  I  I  K  D  R  A
964 gctctggccgacccgtcaaactcagtgcaccagatgtgggagtac
    A  L  A  D  P  S  N  S  V  H  Q  M  W  E  Y
1009 gacagccagacggcgttggtggcccgcgccaagaggtag 1047
     D  S  Q  T  A  L  V  A  R  A  K  R  *
```

**Figure 5.6 The amino acid translation of the cDNA clone (Y-21) of *Gclp1* encoding the full length protein.** This protein has an open reading frame of 312 amino acids. The start and stop codons are in bold red. The putative transit peptide is shown in bold green.


**Figure 5.7 The alignment of Gclp1 with other γ-CAs**. The zinc coordination residues are shown in red. The γ-CA from *Methanosarcina thermophila* is Cam and the γ-CA homologue of *Synechococcus sp.* is CcmM. Cam has an open reading frame of 247 amino acids. The CcmM protein from *Synechococcus* has a long C-terminal extension. CcmM has an open reading frame of 539 amino acids. The Gclp1 protein from *Chlamydomonas reinhardtii* has an open reading frame of 312 amino acids. * represents a completely conserved amino acid, : represent conserved amino acid substitutions, and . represent semi conserved amino acid substitutions.

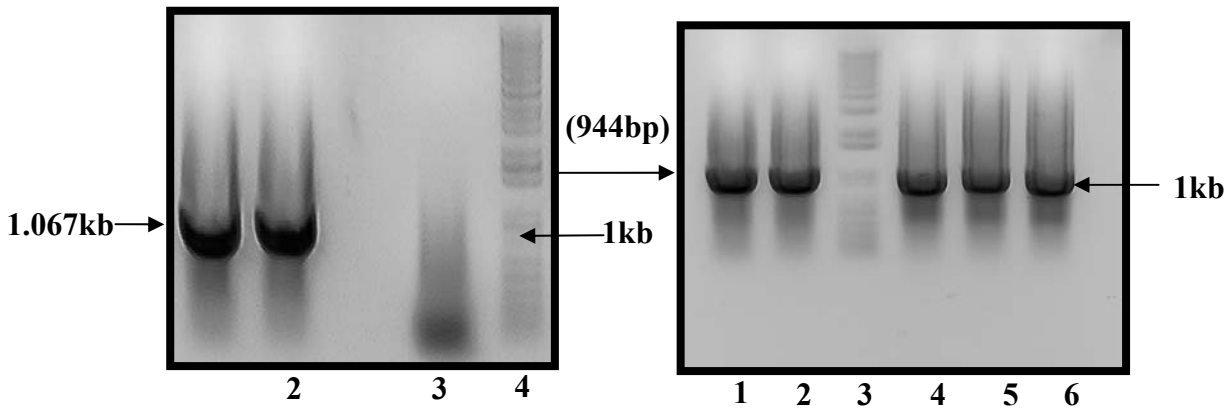



**Figure 5.8 PCR for the construction of the *Gclp1* overexpression plasmid. A.** Lane 1 and Lane 2 show PCR results using the S1 and M1RH primers on the cDNA library. Lane 3 shows PCR result using the same set of primers in the absence of DNA. Lane 4 shows a 1 kb DNA ladder. B. Lanes 1, 2, 4, 5 and 6 show PCR results using S3 and M1RH primers on the cDNA library. Lane 3 shows a 1 kb DNA ladder.

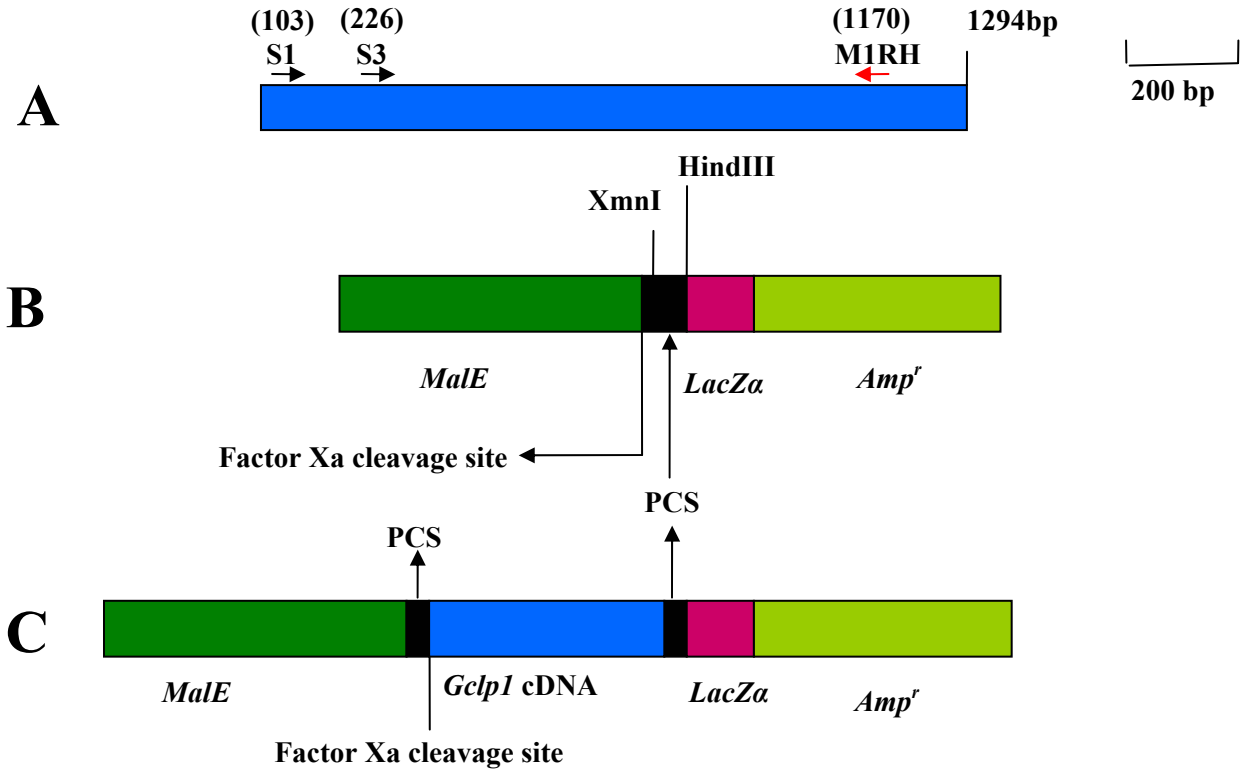



**Figure 5.9 A schematic diagram of the recombinant pMal-Gclp1 expression construct.** A. A schematic figure showing the alignment of primers (S1/S3 and M1RH) used to amplify G*clp1* cDNA. The M1RH primer has a HindIII site incorporated at the 5′end. The numbers on the top denote the primer positions in base pairs. B. Part of the pMal vector showing the polylinker cloning site (PCS), β-galactosidase (*LacZα*) and β-lactamase (*Amp*$^r$) genes, XmnI and HindIII recognition sites and Factor Xa cleavage site. *MalE* codes for the MBP (maltose binding protein). C. A schematic figure of the *Gclp1* recombinant construct.

into the pMal-c2x vector. The latter was done to check for any missense or null mutation that could have been introduced by the DNA polymerase during PCR. Two clones were selected out of the thirty recombinant clones of each type. Clone Y-21 had the cDNA insert coding for the full length Gclp1 protein while clone Y-9 had the cDNA insert coding for the putative mature Gclp1 protein.

**4. Overexpression and Purification of the Recombinant MBP-*Gclp1* Fusion Protein**

Both the clones were tested for overexpression of Gclp1. Recombinant *E. coli* cells were induced with 1 mM IPTG for two hours at 37°C for overexpression of the fusion protein. Equal concentrations of protein from induced and uninduced *E. coli* cells were loaded on a 12% SDS-polyacryamide gel (Fig. 5.11). The overexpression of the recombinant full length Gclp1 was 30%-40% lower than that of Cah3 and Cah6 under identical inducing conditions. The overexpressed recombinant Gclp1 proteins were purified by affinity chromatography using amylose resin according to a protocol given in the NEB technical bulletin. The fusion proteins were cleaved by the serine protease, Factor Xa, for four hours at 23°C to separate Gclp1 from the MBP. Purification and cleavage of fusion protein was confirmed by performing SDS-PAGE (Fig. 5.12). Moreover, the Gclp1 protein was very susceptible to proteolytic degradation during purification. SDS-PAGE results showed that Factor Xa cleaved fusion protein (from Y-21 clone) specifically to yield MBP (42 kDa) and Gclp1 (36 kDa). Along with this specific cleavage, there was evidence of nonspecific cleavage of Gclp1 fusion protein yielding a fragment of 50 kDa and a fragment of 28 kDa. This nonspecific cleavage was enhanced in the presence of Factor Xa and was not prevented by the presence of the common protease inhibitors.

**5. CA Activity Assays and Immunoblotting Results**

No CA activity was detected in the crude cell extracts or Factor Xa cut and uncut purified fusion protein from either type of recombinant Gclp1 clone. Under the same CA activity assay

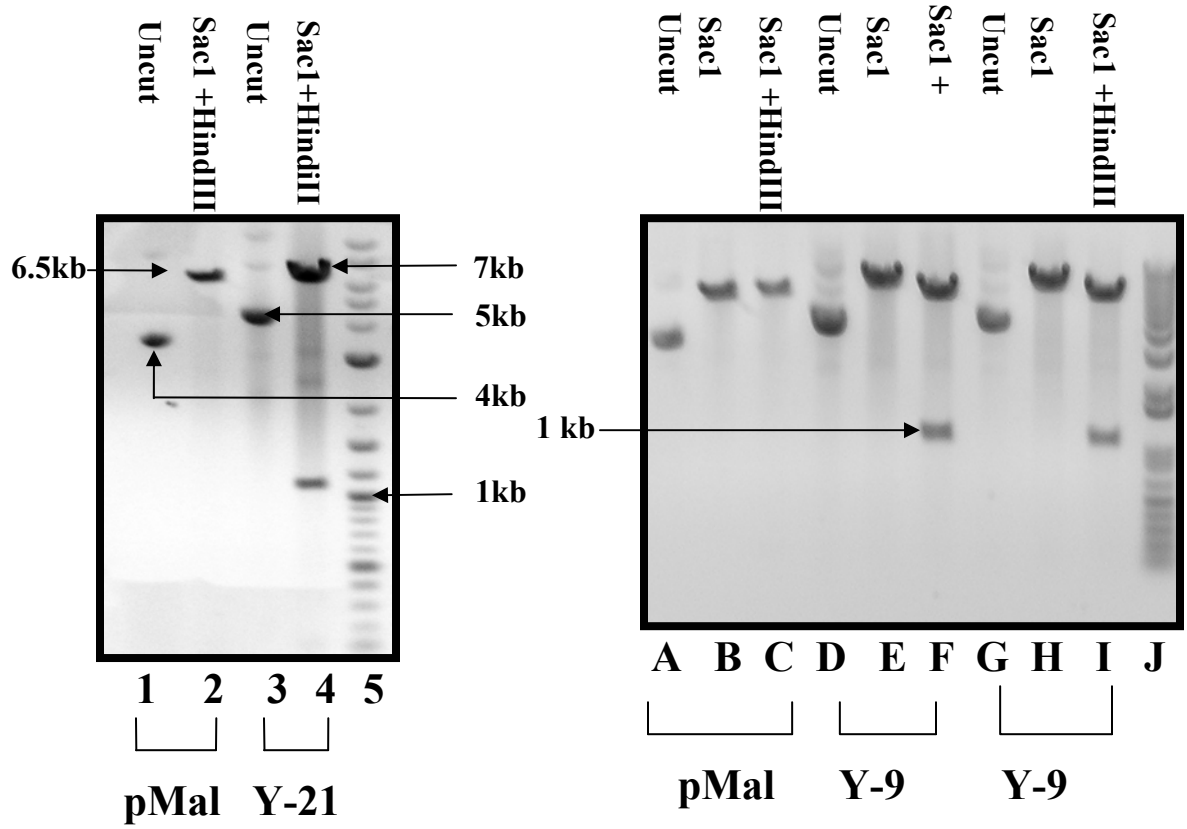



**Figure 5.10 The restriction enzyme digestions of constructs Y-21, Y-9 and the pMal vector to verify cloning of *Gclp1*.** Lanes 1 to 4 represent restriction digestions of Y-21 and pMal plasmids with SacI and HindIII. Lane 5 contains a 2-log DNA ladder (NEB). Lanes A to I represent restriction digestions of Y-9 and the pMal plasmids with SacI and HindIII. Lane J contains a 1 kb DNA ladder (NEB). Double digests of Y-21 and Y-9 with SacI and HindIII yield the DNA inserts of size 1.1 kb and 1 kb.

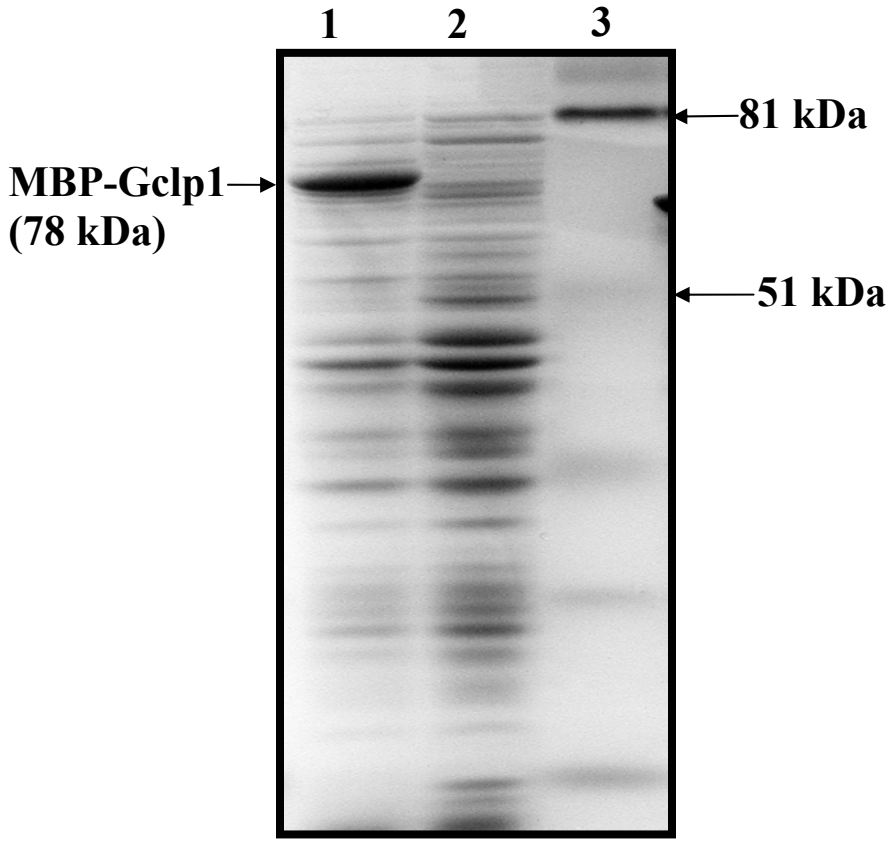



**Figure 5.11 A 12% SDS-polyacrylamide gel showing the uninduced and induced *Escherichia coli* cells expressing the recombinant MBP-Gclp1 protein.** Lane 1 and Lane 2 contain 30 µg of proteins from induced and uninduced *E. coli* cells, respectively. Lane 3 contains prestained molecular weight markers.

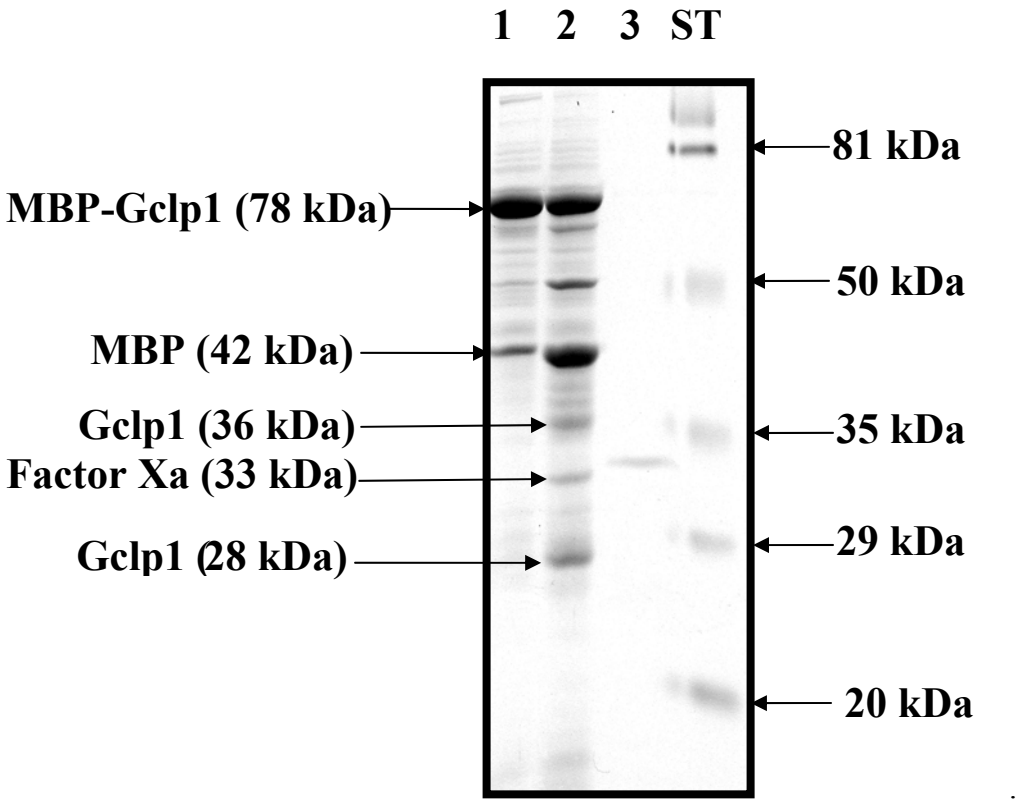



**Figure 5.12 A 12% SDS-polyacrylamide gel showing the purified uncut and Factor Xa cut MBP-*Gclp1* protein.** Lane ST represents the prestained molecular weight markers. Lane 1 contains 20 μg of uncut chimeric MBP-Gclp1. Lane 2 contains 30 μg of chimeric MBP-*Gclp1* cut by 1 μg of Factor Xa. Lane 3 contains 1 μg of Factor Xa. Proteins were stained with Coomassie Blue after 12% SDS-PAGE.

conditions, a crude extract of *E. coli* cells overexpressing Cam (γ-CA of *M. thermophila*) and the purified recombinant Cam showed significant amounts of CA activities with specific activities of about 0.30 WAU/mg and 800 WAU/mg, respectively. Crude recombinant *E. coli* cells overexpressing Gclp1 and Factor Xa cut and uncut purified recombinant Gclp1 fusion proteins were probed by the primary antibody raised against the γ-CA protein of *M. thermophila* in Western blotting. The immunoblotting showed that the antibody raised against the archaebacterial γ-CA protein did not cross react with Gclp1 (data not shown).

**DISCUSSION**

Here I report the identification of a putative γ-CA gene from *C. reinhardtii* for the first time. Since its discovery in the archaebacterium *Methanosarcina thermophila*, several putative γ-CA genes have been identified in eubacteria and plants. In the methanogen *M. thermophila*, a metabolic switch from methanol to acetate utilization results in the elevation of CA activity suggesting this enzyme is important for growth on acetate. It has been proposed that the CA might be required for a $CH_3CO_2^-/H^+$ symport system or for efficient removal of cytoplasmically produced $CO_2$ (Alber and Ferry, 1994). A γ-CA homologue, CcmM, was discovered earlier in *Synechococcus* PCC7942 (Price *et al.,* 1993). CcmM does not have any detectable CA activity (Moroney *et al*., 2001) but is clearly required for optimal growth on low $CO_2$ (Price *et al.,* 1993).

The homology search analysis using protein prediction programs (P SORT, CHLOR P and TARGET P listed under http://ca.expasy.org/tools/#translate) show that Gclp1 is either located in the cytoplasm or is secreted outside the plasma membrane. It has three histidine residues that could serve as zinc coordination residues in the amino acid sequence, as seen in Cam. Gclp1 aligns well with CcmM and the *M. thermophila* γ-CA, Cam. The full length and mature Gclp1 have a calculated molecular weight of 28 kDa and 34 kDa, respectively. The calculated pI of the full length and mature Gclp1 are 6.54 and 6.36, respectively.

No detectable CA activity like that found in the uncut or Factor Xa cut purified recombinant Cah3 and Cah6 fusion proteins has been found in the heterologously produced recombinant Gclp1 fusion protein. Under the same CA activity assay conditions, crude extracts of *E. coli* cells overexpressing the γ-CA of *M. thermophila* (Cam) and purified Cam showed significant amount of CA activity. Although this seems to indicate that Gclp1 might be an enzymatically inactive CA in *C. reinhardtii*, there can be other explanations for the inactivity of the enzyme. For example,CcmM, the γ- CA homologue from *Synechococcus*, lacks detectable CA activity (Moroney *et al*., 2001) but it is required for growth under low $CO_2$ conditions. Comparisons of SDS-PAGE results of overexpression and purification of recombinant Cah3, Cah6 and Gclp1 fusion proteins indicate that the overexpression of recombinant Gclp1 was 30%-40% lower than that of Cah3 and Cah6 under identical inducing conditions. Moreover, the Gclp1 protein was susceptible to proteolytic degradation during purification (Fig. 5.12). This nonspecific cleavage was enhanced in the presence of factor Xa and was not prevented by the presence of common protease inhibitors. The above results suggest the possibilities that the Gclp1 protein is degraded as a response to possible toxicity to the *E. coli* cells, or that the chimeric protein is unstable.

On the other hand, it is also possible that the Gclp1 has become enzymatically inactive in the course of evolution. Gclp1 has 40%-43% amino acid identity with an APF1 transcription factor, a putative ferripyochelin binding protein in *Arabidopsis* and also to some proteins of unknown function in *Oryza sativa* and *Zea mays*. Hence, other functional roles of Gclp1 *in vivo* cannot be ruled out.

The *Chlamydomonas* EST database has another putative γ-CA homologue, *Gclp2* which is 80% similar to *Gclp1*. The *Arabidopsis* genome contains at least three homologues of γ-CA; but it is not yet known if these gene products are active CAs or what physiological role they

might be playing in the cell. *Gclp1* and *Gclp2* have more than 40 ESTs. Northern blot analyses of these two genes using air and high $CO_2$ adapted cells can be used in future research to indicate if the expression of these genes is regulated by the $CO_2$ concentration in the air. Also, generation of mutants of these putative γ-CA genes by antisense RNA or RNAi might allow a better understanding of the functional roles of these genes.

# CHAPTER 6

# CONCLUSIONS

There has been a surge of interest in CAs from plants and algae over the past decade. This interest began with the discovery of the β-CA in plants in 1990 (Fawcett *et al*., 1990) and has continued with the finding of multiple α- and β-CAs in *Chlamydomonas reinhardtii* and *Arabidopsis thaliana* and the determination of the critical physiological roles, CAs have in cyanobacteria and macro-algae. Analysis of the *Arabidopsis* genomic database reveals at least 14 genes, potentially encoding CAs, that have homologies with ESTs and are expressed in cells (Moroney *et al*., 2001). Clearly the number of CAs in plants is much greater than previously thought. *C. reinhardtii*, a unicellular green alga, is not far behind with five CAs already identified (Table 1.2; Moroney *et al*., 2001).

The primary goals of my work were to a) identify, clone and overexpress novel CA and CA-like genes from the green unicellular alga *Chlamydomonas reinhardtii*, b) partially characterize recombinant CA proteins if they are active, and c) determine if the new CAs play any role in photosynthesis and CCM of the green alga. In addition to characterizing the recombinant Cah3 protein encoded by the *Cah3* gene, identified by Karlsson *et al*., (1995), I have identified one novel β-CA gene (*Cah6*), one new putative β-CA gene (*Cah7*) and two putative γ-CA like genes (*Gclp1 and Gclp2*) in *C. reinhardtii* (Table 6.1). This brings the total number of CA/CA-like proteins in *C. reinhardtii* to nine (Table 6.1). I have concentrated on *Cah6* and *Gclp1* for the research described herein.

Karlsson *et al*. (1998) published the sequence of the *Cah3* cDNA. Although the *Cah3* gene has been well characterized, there is no report of the activity of the purified protein. This is

**Table 6.1** A summary of CA and CA like genes known to date in *Chlamydomonas reinhardtii*

| Gene family | Known gene/s | | References |
|---|---|---|---|
| | Gene name | Location | |
| α | *Cah1*, *Cah2* | periplasm | Fukuzawa *et al*. 1990; Fujiwara *et al*.1990 |
| α | *Cah3* | thylakoid | Karlsson *et al*. 1995 |
| β | *Ca1, Ca2* | mitochondria | Eriksson *et al*. 1996 |
| β | ***Cah6*** | chloroplast** | This dissertation |
| β | ***Cah7*** | unknown | Gene model predictions and EST database |
| γ | ***Gclp1*** | cytoplasm* | This dissertation |
| γ | ***Gclp2*** | cytoplasm * | Gene model predictions and EST database |

* based on analyses by protein prediction programs PSORT, TARGET P and CHLOR P; ** based on immunolocalization results. The names of CA genes identified after 1998 are given in bold.

the first report of detection of significant CA activity in the purified Cah3 protein and partial characterization of this CA activity. α-CAs are known to have a broad range of specific activity ranging from 5 WAU/mg (human CAIII) to 2000 WAU/mg (human CAII). The specific activity of the recombinant Cah3 enzyme, which is around 735 WAU/mg, falls within this range.

Recombinant Cah3 was purified to raise a "new" Cah3 antibody that would be a more dependable tool for Western blotting and immunolocalization experiments than the existing "old" Cah3 antibody. Western blotting results using the new Cah3 antibody have shown that the protein is missing in the two allelic *cah3* mutants, *ca-1* and *cia3* (Fig. 3.6, Chapter 3). The immunoblotting results also indicate that the "new" Cah3 antibody specifically detects the Cah3 and is more dependable as a tool in performing immunolocalization studies. The Cah3 enzyme requires maintenance in the oxidized state to retain its CA activity, as indicated by a reduction of activity following exposure to thiol reducing agents. This seems to indicate that disulfide bond formation in Cah3 is essential.

Researchers in several laboratories have tried to assay CA activity in chloroplasts. Using mass-spectrometric measurements of $^{18}O$ exchange, Sültemeyer *et al*., (1995) have characterized two chloroplastic CA activities in *C. reinhardtii* cells. One CA activity is an "insoluble" form associated with the thylakoid fraction and is more insensitive to EZ while the other is a "soluble" form and sensitive to EZ (Amoroso *et al*., 1996). Villarejo *et al*., (2001) have found a new chloroplast envelope CA activity that is sensitive to EZ and is constitutively expressed. Support for the presence of the insoluble chloroplastic CA activity is provided by the identification of the 29 kDa Cah3 which is located on the luminal side of thylakoid membrane (Karlsson *et al*., 1998). I have found Cah3 to be sensitive to EZ, contrary to the finding of Sültemeyer *et al*., (1995).

My research data show that Cah6, the sixth carbonic anhydrase gene and the third β-CA gene to be identified in *C. reinhardtii*, is located in the stroma of the chloroplast and can serve as

the soluble chloroplastic CA but it is less sensitive to EZ contrary to the findings of Sültemeyer *et al*., (1995). Cah6 has about 20 ESTs in the data base at this time. Cah6 is less susceptible to sulfonamide inhibitors, a characteristic trait of all active β-CAs. β-CAs in higher plants are known to have specific activities ranging from 600 WAU/mg to 1000 WAU/mg. The specific activity of the recombinant Cah6 protein is 940 WAU/mg and falls within the activity range of β-CAs.

Both Cah3 and Cah6 are constitutive in expression but are slightly upregulated under low $CO_2$ conditions (Karlsson *et al*., 1998; Chapter 4 of this dissertation, 2003). Another mutant of *C. reinhardtii*, *cia5,* has a HCR phenotype, fails to induce CCM and does not synthesize any of the low $CO_2$ inducible mitochondrial and periplasmic CAs along with other low $CO_2$ inducible polypeptides. It is deficient in a putative transcription factor coded by the *Cia5* gene (Fukuzawa *et al*., 2001; Xiang, *et al*, 2001). Katzman *et al*., (1994) used a $^{14}C$ assay to measure the CA activity in the chloroplasts of *cia5* and wild type cells of *Chlamydomonas* and found that the CA activity of *cia5* is almost identical to that in wild type cells. The "new" Cah3 and Cah6 antibody detects the Cah3 and Cah6 protein, respectively in the *cia5* cells (data not shown). So my research data on Cah3 and Cah6 agrees with the earlier work of Katzman *et al*., (1994). This also indicates that probably Cah3 and Cah6 gene expressions are not under the regulation of the *Cia5* gene.

Comparative analyses of characteristic properties of Cah3 and Cah6 CA activities are shown in (Table 6.2). The sulfonamide $I_{50}$ of Cah6 was 100- to 1000- fold higher than that of Cah3, a characteristic trait of all β-CAs. The optimum temperatures for CA activity of Cah6 and Cah3 were around 32°C-33°C and 50% of the catalytic activities of these proteins were lost above 43°C (Table 6.2). Thus the optimal temperature for Cah6 activity, like that of Cah3, falls within the range of the optimal growth temperature (12°C- 35°C) of *C. reinhardtii*. β-CAs

localized to higher plant chloroplasts have been reported to be sensitive to oxidation and therefore are dependent on a reducing environment to retain their catalytic activity (Tobin, 1970; Atkins *et al*., 1972; Cybulsky *et al*., 1979; Johansson and Forsman, 1993). In contrast, the activity of Cah6 like that of the cytosolic β-CA of the green alga *Coccomyxa* was not affected by thiol reducing agents (Table 6.2). Cah6 also has 34% sequence identity with the cytoplasmic β-CA of *Coccomyxa*.

Immunolocalization results show that Cah6 is localized in the stroma of the chloroplast and is not present in the pyrenoid. Interestingly, it is four fold more abundant in the area around the pyrenoid, particularly in the starch sheath surrounding the pyrenoid than in the other areas in the stroma of the chloroplast. My hypothesis is that Cah6 might indirectly play a role in CCM by trapping any $CO_2$ leaking out from the pyrenoid (the site of localization of Rubisco in *C. reinhardtii*) into $HCO_3^-$. This would help to maintain the $HCO_3^-$ pool in the stroma, which is essential for the operation of CCM.

RNA interference was used to generate Cah6 mutants to provide insight into the functional role of Cah6 in the chloroplast of *C. reinhardtii*. None of the RNAi transformants screened so far show any visible difference in the expression of Cah6 protein in the mutant cells compared to that in the wild type cells. Hence no deduction can be made about the functional role played by Cah6 in the CCM and photosynthesis based on this approach. Further screening of the air adapted cells of remaining transformants by Western blotting will provide definite indication as to whether or not RNA interference is working in the wild type *C. reinhardtii* cells. It might be possible even if the Western blotting results show that RNAi is working in these transformants, Cah6 mutants might not show any phenotypic difference from the wild type cells if it is playing an indirect role in CCM. Specific reduction of chloroplast CA activity by antisense RNA in transgenic tobacco plants had a minor effect on photosynthetic $CO_2$

assimilation (Price *et al*., 1994). Plants with even the lowest CA levels (2% of wild type levels) were not morphologically distinct from the wild type plants (Price *et al*., 1994). CA has a high turnover number and is one of the most catalytically efficient enzymes known to date. Hence it might be necessary to underexpress a CA gene to a level of 99%-100% in any RNAi CA mutant to see a visible phenotypic difference between the wild type and mutant cells.

Since its discovery in the archaebacterium *Methanosarcina thermophila* (Alber and Ferry, 1994), several putative γ-CA genes have been identified in eubacteria and plants. I have identified a putative γ-CA gene (*Gclp1*) in *C. reinhardtii* for the first time. The homology search and the multiple sequence analysis show that Gclp1 has significant homology to the one known active γ-CA (Cam) from *M. thermophila* and the putative γ-CA, CcmM from *Synechococcus*. It has three histidine residues that can function as zinc coordination residues in the amino acid sequence, a characteristic of γ-CAs. Gclp1 was cloned in the overexpression vector pMal-c2x. Like *Cah6*, two different expression constructs were made. One clone contained the ORF of Gclp1 and the other contained the cDNA sequence that would code for the mature Gclp1 protein. No detectable CA activity was found in the heterologously produced recombinant Gclp1 fusion protein from either clone. Under the same CA activity assay conditions, crude extracts of *E. coli* cells overexpressing Cam CA activity CA showed significant amount of CA activity. If the overexpression of the Gclp1 is toxic to *E. coli* cells, bacteria can die or can degrade the overexpressed protein or specifically introduce mutations in the DNA sequence coding for the toxic protein (personal communication, Dr. S.G. Bartlett, Louisiana State University, Baton Rouge, Louisiana). If expression of Gclp1 was toxic to bacteria, clones containing Gclp1 were lost and never selected. Alternatively, it can also be possible that the cells were overexpressing mutated Gclp1 or degrading the Gclp1 that they were expressing. Gclp1 was also highly susceptible to degradation. Any one of these could explain the lack of detectable activity.

There are quite a few questions that still remain to be answered to shed light on the intracellular locations and functional roles of these new CAs and the biochemical characteristics of CA activities. Analyses of Cah3 using various post translational prediction programs (PSORT, CHLOR P and TARGET P) suggest that the protein is located in the thylakoid. Thermolysin (a protease)-treated intact and sonicated thylakoids from wild type and *cah3* mutant *cia3* were analyzed by Western blotting using the "old" cah3 antibody (Karlsson *et al*., 1998). Previously published results using the "old" antibody showed that Cah3 is present on the luminal side of thylakoids and the mutant *cia3* has the protein (Karlsson *et al*., 1998). Due to the lack of dependability of the "old" antibody, this experiment should be repeated with the "new" antibody. We are currently performing immunolocalization of Cah3 in WT137, *cia3* and *ca-1* cells, using this "new" antibody. The immunolocalization results along with the Western blotting analyses of thermolysin-treated intact and sonicated thylakoids from WT 137, *ca-1* and *cia3* will be helpful in reaching a definite conclusion regarding the intracellular location of Cah3 and its presence or absence in the two *cah3* allelic mutants.

Cah3 has three cysteine residues in the mature protein. In the future, site directed mutagenesis experiments can be done to specifically substitute these cysteines in Cah3 in order to determine if these cysteines play a role in controlling the CA activity of Cah3, as they do in other α-CAs.

Furthermore, Cah3 (generated by cleavage of MBP-Cah3 fusion protein by Factor Xa and separated from MBP by SDS-PAGE) can be eluted from the polyacrylamide gel sections and the free thiol content of oxidized and reduced Cah3 enzyme estimated from the increase in absorbance ($A_{412}$) caused by formation of a 2-nitro-5 thiobenzoate anion resulting from cleavage of DTNB [5′, 5′-dithiobis (2-nitrobenzoic azide)] upon reaction with a thiolate anion (Hiltonen *et al*., 1998).

**Table 6.2** The comparison of some characteristic properties of the recombinant Cah3 and Cah6 activity

| Biochemical traits | Mature Cah3 | Cah6 (ORF) |
|---|---|---|
| Specific activity | 730 WAU/mg | 940 WAU/mg |
| Apparent molecular weight | 29 kDa | 31 kDa |
| pI | 7.87 | 7.0 |
| Thermostability | Loses 50% of CA activity at 43°C ± 0.7°C | Loses 50% of CA activity at 43°C ± 0.8°C |
| Effect of SH-reducing agent | Inhibits CA activity by 50% ± 1.4% to 82% ± 1.4% | No effect on CA activity |
| Sulfonamide Inhibition | EZ $I_{50}$ ($6 \times 10^{-9} \pm 0.7 \times 10^{-9}$ to $8 \times 10^{-9} \pm 0.9 \times 10^{-9}$M) | EZ $I_{50}$ ($2 \times 10^{-6} \pm 0.3 \times 10^{-6}$ to $9 \times 10^{-6} \pm 0.5 \times 10^{-6}$ M) |
| Azide Inhibition | $I_{50}$ ($3.2 \times 10^{-5} \pm 0.2 \times 10^{-5}$ M) | $I_{50}$ ($1.5 \times 10^{-5} \pm 0.4 \times 10^{-5}$ M) |
| Cyanide Inhibition | $I_{50}$ ($5.9 \times 10^{-5} \pm 0.6 \times 10^{-5}$ M) | $I_{50}$ ($5 \times 10^{-6} \pm 0.5 \times 10^{-6}$ M) |

The data shown in the table for thermostability, effect of SH-reducing agents, sulfonamide, azide and cyanide inhibition are the averages ± SD of three different measurements. The specific activity calculated is calculated taking into account the molecular size ratios of Cah3 and Cah6 and MBP.

Some reasons for the CA encoded by the B3 clone not having CA activity are discussed in Chapter 4 of this dissertation. Resequencing of the Cah6 gene in the B3 clone needs to be done to check for any introduced mutations by the bacteria cells, before arriving at a conclusion about the activity of the purified Cah6 protein from B3 clone. If no mutations are found in the Cah6 sequence in the B3 clone, then site directed mutagenesis experiments are needed to alter the arginine residue in the enzymatically active Cah6 from B48 clone to study if the single arginine residue in Cah6 plays a role in modulating the catalytic activity of Cah6. The Cah6 protein has two cysteines and one histidine residue that can act as zinc coordinating ligands, as is seen in enzymatically active β-CAs. These conserved residues can be altered by site directed mutagenesis to verify their role in modulating the catalytic activity of Cah6.

Recently, a thorough search of the genomic and EST database of *C. reinhardtii* led to the identification of another putative β-CA gene. We have named it *Cah7*. Cah7 has only one EST record. Hence it is expressed in the cell but at low level. BLAST searches with Cah7 protein sequence showed that Cah7 has 28%-33% sequence identity with the mitochondrial Ca1 and the chloroplastic Cah6 protein from *C. reinhardtii*. It has 50% homology with the cytoplasmic β-CA protein from the green alga *Coccomyxa*. The 5' end of this gene needs to be clarified by rapid amplification of cDNA ends (RACE) before one can make any prediction about the localization of Cah7. Once the 5′ end of *Cah7* is resolved, it can be overexpressed in the pMal-c2x vector to study its biochemical properties and to raise an antibody against it, if the protein is enzymatically active. If Cah7 is found active, gene specific mutants can be generated using antisense or RNA interference technique to study its functional role, if any, in photosynthesis and CCM.

Activities of Ca1, Ca2, Cah3, Cah7 (cellular location still unknown) and Cah6 jointly constitute the total intracellular CA activity. Ca1 and Ca2 (mitochondrial CAs) have only one amino acid difference in the transit peptide. Hence the mature mitochondrial β-CAs (Ca1 and

Ca2) are identical in amino acid sequence. Ca1 has been cloned in the pMal-c2x vector to be expressed as a MBP-fusion protein (Eriksson *et al*., 1994). There is no report on the measurement of the CA activity in the purified recombinant fusion Ca1. In future experiments the recombinant Ca1 can be purified and the CA activity can be biochemically characterized. Cyanobacterial β-CA and the "soluble" chloroplast CA activity of *C. reinhardtii* have been shown to possess a high requirement for $Mg^{2++}$ (Price *et al*., 1992; Amoroso *et al*., 1996). The effect of $MgSO_4$, different CA inhibitors and thiol reducing agents, on the activities of soluble known CAs (Cah6, Cah7, Ca1 and Ca2) can be studied in future. The biochemical characteristics (particularly $I_{50}$ of different CA inhibitors and the effect of $Mg^{2++}$ on the CA activity) of Ca1/Ca2, Cah7, Cah3 and Cah6 can be exploited to selectively study the CA activity of each of the intracellular CAs *in vivo*.

Analyses of the expression patterns of CAs known to be involved in CCM like *Cah3* and *Cah1* in the Cah6 RNAi mutants by Northern blotting can be helpful in determining if there is a relative difference in the expression patterns of these CAs in the Cah6 mutants and the wild type cells. This experiment can give us some indication, if in the Cah6 RNAi mutants the expression of the CCM involved CAs, are upregulated to compensate for the underexpression of Cah6. Immunoblotting and Northern blotting analysis (using Cah3 and Cah6 specific probe) of *cia5* and wild type cells of *C. reinhardtii* can show the expression patterns of the *Cah3* and *Cah6* genes in this mutant and can indicate if these genes are under the regulation of the *Cia5* gene.

Maintenance of the $HCO_3^-$ pool in the stroma is essential for the operation of CCM. To prevent the dissipation of the $HCO_3^-$ pool, CA activity is absent in the cytosol of cyanobacteria. The expression of human CA in the cytoplasm of *Synechococcus* PCC7942 cells results in a massive leakage of $CO_2$ from the cells producing a HCR phenotype and implicating the carboxysomes as the site of $CO_2$ elevation (Price and Badger, 1989a). Analyses of signal peptide

seem to indicate that Gclp1 is located in the cytoplasm (Chapter 5). *Coccomyxa*, a green alga does not have a CCM (Palmqvist *et al*., 1994). It has a cytoplasmic β-CA. It might be possible that the Gclp1 has become catalytically inactive in the course of evolution of CCM in *Chlamydomonas* or has some other enzymatic roles. Interrelationship between the presence or absence of CCM and cytoplasmic CAs in phylogenetically related green algae can be studied to arrive at an answer. Gclp1 has 40%-43% homology to some unknown proteins in rice, maize and *Arabidopsis* and also to ferripyochelin binding proteins in bacteria. Hence other physiological roles of Gclp1 cannot be ruled out.

Certain proteins that are toxic to bacteria can be overexpressed in yeast cells. Gclp1 can be overexpressed in eukaryotic yeast cells to check for CA activity. An antibody can be generated against the purified Gclp1 protein to immunolocalize Gclp1 in *Chlamydomonas* cells. This can verify its location in cytoplasm. The *Chlamydomonas* EST database has another putative γ-CA homologue, Gclp2 which is 80% similar to Gclp1. The *Arabidopsis* genome contains at least three homologues of γ-CA but it is not yet known if these gene products are active CAs or what physiological roles they might be playing in the cell. Gclp1 and Gclp2 have more than 40 ESTs and are highly expressed in cells. Northern blot analyses of these two genes using air and high $CO_2$ adapted cells can indicate if the expression of these genes is regulated by $CO_2$ concentration in the air. Since Gclp1 has similarities with ferripyochelin binding proteins, Northern blot analyses using cells grown in the presence and absence of iron in the nutrient medium, can indicate if iron concentration in the external medium can control the expression of these genes. In the future, generation of mutants of these putative γ-CA genes by antisense RNA or RNAi might be useful in probing the functional roles of these genes.

To our knowledge no other β- α- or γ-CA genes, other than those that have been already identified (Table 6.1), has been found in the EST and genomic database of *C. reinhardtii*. At this

point, *Arabidopsis* has five, six and three different genes that align well with the α-, β- and the γ-CAs, respectively. The availability of *Arabidopsis* and *Chlamydomonas* genome sequences and EST databases can be used to find out the exact number of expressed CA isoforms in these organisms. The challenge for future researchers will be to determine the expression patterns, localization and physiological roles for each of these isoforms. As there appears to be a large number of isoforms in plants and algae, CA researchers should be busy in the near future.

# REFERENCES


Aizawa K. and Miyachi S. (1986) Carbonic anhydrase and $CO_2$ concentrating mechanisms in microalgae and cyanobacteria. FEMS Microbiol. Rev. **39:** 215- 233.

Akita S. and Tanaka I. (1973) Studies on the mechanism of differences in photosynthesis among species. IV. The differential response in dry matter production between C3 and C4 species to atmospheric carbon dioxide enrichment. Proc. Crop. Sci. Soc. Jpn. **42:** 288-295.

Alber B.E. and Ferry J.G. (1994) A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. Proc. Natl. Acad. Sci.USA **91:** 6909-6913.

Alber B.E. and Ferry J.G. (1999) A plant type (β-class) carbonic anhydrase in the thermophilic methanoarchaeon *Methanobacterium thermoautotrophicum*. J. Bact. **181:** 6247- 6253.

Alber B.E., Colangelo C.M., Dong J., Stalhandske C.M.V., Baird T.T., Tu C., Fierke C.A., Silverman D.N., Scott R.A. and Ferry J.G. (1999) Kinetic and spectroscopic characterization of gamma carbonic anhydrase from the methanoarcheon *Methanosarcina thermophila*. Biochemistry, **38:** 13119- 13128.

Altschul S.F., Madden T.L., Schäffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl. Acids Res. **25:** 3389-3402.

Amoroso G., Sültemeyer D.F., Thyssen C. and Fock H.P. (1998) Uptake of $HCO_3^-$ and $CO_2$ in cells and chloroplasts from the microalgae *Chlamydomonas reinhardtii* and *Dunaliella tertiolecta*. Plant Physiol. **116:** 193-201.

Amoroso G., Weber C., Sültemeyer D.F. and Fock H.P. (1996) Intracellular carbonic anhydrase activities in *Dunaliella tertiolecta* (Butcher) and *Chlamydomonas reinhardtii* (Danegeard) in relation to inorganic carbon concentration during growth: further evidence for the existence of two distinct carbonic anhydrases associated with the chloroplast. Planta, **199:** 177-184.

Andrews T.J. and Lorimer G.H. (1987) Rubisco: structure, mechanisms and prospects for improvement. In: The Biochemistry of Plants, editors Hatch, M.D. and Boardman, N.K., Academic Press, New York, NY, USA.

Arnon D.I. (1949) Copper enzymes in isolated chloroplasts; Polypeptide in *Beta vulgaris*. Plant Physiol. **66:** 407-413.

Atkins C.A., Patterson B.D. and Graham D. (1972) Plant carbonic anhydrases II. Preparation and some properties of monocotyledon and dicotyledon enzyme types. Plant Physiol. **50:** 218-223.

Badger M.R. (1987) The $CO_2$ concentrating mechanism in aquatic phototrophs. In: The Biochemistry of Plants. A comprehensive treatise. Academic Press, San Diego, CA, USA. Vol. **10:** 219-274.

Badger M.R. (2003) The roles of carbonic anhydrases in photosynthetic $CO_2$ concentrating mechanisms. Photosynth. Res. **77:** 83-94.

Badger M.R. and Pfanz H. (1995) Effect of carbonic anhydrase inhibition on photosynthesis by leaf pieces of $C_3$ and $C_4$ plants. Aust. J. Plant Physiol. **22:** 45-49.

Badger M.R. and Price G.D. (1992) The $CO_2$ concentrating mechanism in cyanobacteria and microalgae. Physiol. Plant. **84:** 606-615.

Badger M.R. and Price G.D. (1994) The role of carbonic anhydrase in photosynthesis. Ann. Rev. Plant Physiol. and Plant Mol Biol. **45:** 369-392.

Badger M.R., Andrews T.J., Whitney S.M., Ludwig M., Yellowlees D.C., Leggat W. and Price G.D. (1998) The diversity and co-evolution of Rubisco, plastids, pyrenoids and chloroplast-based CCMs in the algae. Can. J. Bot. **76:** 1052-1071.

Badger M.R., Hanson D.T., and Price G.D. (2002) Evolution and diversity of $CO_2$ concentrating mechanisms in cyanobacteria. Funct. Plant Biol. **29:** 161-173.

Badger M.R., Kaplan A., and Berry J.A. (1980) Internal inorganic carbon pool of *Chlamydomonas reinhardtii*. Evidence for a carbon dioxide concentrating mechanism. Plant Physiol. **66:** 407-413.

Beardall J. (1981) $CO_2$ accumulation by *Chlorella saccharophila* (Chlorophyceae) at low external pH: evidence for the active transport or inorganic carbon at the chloroplast envelope. J. Phycol. **17:** 371-373.

Bédu S., Pozuelos P., Cami B. and Joset F. (1995) Uptake of inorganic carbon in the cyanobacterium *Synechocystis* PCC 6803: physiological and genetic evidence for a high affinity uptake system. Mol. Microbiol. **18:** 559-568.

Beer S., Israel A., Drechsler Z. and Cohen Y. (1990) Photosynthesis in *Ulva fasciata*. V. Evidence for an inorganic carbon concentrating system and ribulose-1,5-bisphosphate carboxylase/oxygenase kinetics. Plant Physiol. **94:** 1542-1546.

Berman-Frank I., Erez J. and Kaplan A. (1998) Changes in inorganic carbon uptake during the progression of a dinoflagellate bloom in a lake ecosystem. Can. J. Bot. **76:** 1043-1051.

Berry J., Boynton J.E., Kaplan A. and Badger M.R. (1976) Growth and photosynthesis of *Chlamydomonas reinhardtii* as a function of $CO_2$ concentration. Carnegie Inst. Washington Year Book. **75:** 423-432.

Beuf L., Bédu S., Durand M.C. and Joset F. (1994) A protein involved in co-ordinated regulation of inorganic carbon and glucose metabolism in the facultative photoautotrophic cyanobacterium *Synechocystis* PCC 6803. Plant Mol. Biol. **25:** 855-864.

Borkhsenious O.N., Mason C.B. and Moroney J.V. (1998) The intracellular localization of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. Plant Physiol. **116:** 1585- 1591.

Bowes G. and Salvucci M.E. (1989) Plasticity in photosynthetic carbon metabolism of submerged aquatic macrophytes. Aquat. Bot. **34:** 233-266.

Bracey M.H., Christiansen J., Tovar P., Cramer S.P. and Bartlett S.G. (1994) Spinach carbonic anhydrase: Investigation of the zinc-binding ligands by site-directed mutagenesis, elemental analysis and EXAFS. Biochem. **33:** 13126-13131.

Burnell J.N., Gibbs M.J. and Mason J.G. (1990) Spinach chloroplastic carbonic anhydrase-nucleotide sequence analysis of cDNA. Plant Physiol. **92:** 37-40.

Burnell J.N., Ludwig M. and Sugiyama T. (1999) Accession numbers T02079 and T02080.

Burns B.D. and Beardall J. (1987) Utilization of inorganic carbon by marine microalgae. J. Exp. Mar. Biol. Ecol. **107:** 75-86.

Chen Z.Y., Lavigne L.L., Mason C.B. and Moroney J.V. (1996) A low $CO_2$ inducible gene encoding an alanine: α-ketoglutarate aminotransferase in *Chlamydomonas reinhardtii*. Plant Physiol. **112:** 677-684.

Chirica L., Elleby B. and Lindskog S. (2001) Cloning, expression and some properties of α-carbonic anhydrase from *Helicobacter pylori*. Biochim. Biophys. Acta. **1544:** 55-63.

Coba de la Pena T.C., Frugier F., McKhann H.I., Bauer P., Brown S., Kondorosi A. and Crespi M. (1997) A carbonic anhydrase gene is induced in the nodule primordium and its cell-specific expression is controlled by the presence of *Rhizobium* during development. Plant J. **11:** 407-420.

Coleman J.R. and Grossman A.R. (1984) Biosynthesis of carbonic anhydrase in *Chlamydomonas reinhardtii* during adaptation to low $CO_2$. Proc. Natl. Acad. Sci. USA **81**: 6049-6053.

Coleman B. and Rotatore C. (1995) Photosynthetic inorganic uptake and accumulation in two marine diatoms. Plant cell. Environ. **18:** 919-924.

Cox E.H., McLendon G.L., Morel F.M.M., Lane T.W., Prince R.C., Pickering I.J. and George G.N. (2000) The active site structure of *Thalassiosira weisflogii* carbonic anhydrase 1. Biochem. **39:** 12128- 12130.

Cybulsky D.L., Nagy A. Kandel S.I., Kandel M. and Allen G.G. (1979) Carbonic anhydrase from spinach leaves: chemical modification and affinity labeling. J. Biol. Chem. **254:** 2032-2039.

Elleby B.; Chirica L.C., Tu C., Zeppezauer M. and Lindskog S. (2001) Characterization of carbonic anhydrase from *Neisseria gonorrhoeae*. Eur. J. Biochem. **268:** 1613-1619.

Eriksson M., Karlsson J., Ramazanov Z., Gardeström P. and Samuelsson G. (1996) Discovery of an algal mitochondrial carbonic anhydrase: molecular cloning and characterization of a low-$CO_2$-induced polypeptide in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **93:** 12031-12034.

Eriksson M., Villand P., Gardeström P. and Samuelsson G. (1998) Induction and regulation of expression of a low-$CO_2$-induced mitochondrial carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Physiol. **116:** 637-641.

Espie G.S. and Kandaswamy R.A. (1994) Monesin inhibition of $Na^+$-dependent $HCO_3^-$ transport distinguishes it from $Na^+$ independent $HCO_3^-$ transport and provides evidence for $Na^+/HCO_3^-$ symport in the cyanobacterium *Synechococcus* UTEX 625. Plant Physiol. **104:** 1419-1428.

Fawcett T.W., Browse J.A., Volokita M. and Bartlett S.G. (1990) Spinach carbonic-anhydrase primary structure deduced from the sequence of a cDNA clone. J. Biol. Chem. **265:** 5414-5417.

Fett J.P. and Coleman J.R. (1994) Characterization and expression of two cDNAs encoding carbonic-anhydrase in *Arabidopsis thaliana*. Plant Physiol. **105:** 707-713.

Fisher M., Gokhman I., Pick U. and Zamir A. (1996) A salt-resistant plasma membrane carbonic anhydrase is induced by salt in *Dunaliella salina*. J. Biol. Chem. **271:** 17718-17723.

Friedberg D., Jager K.M., Kessel M., Silman N.J. and Bergman B. (1993) Rubisco but not Rubisco activase is clustered in the carboxysomes of the cyanobacterium *Synechococcus* PCC 7942: Mud-induced carboxysomeless mutants. Mol. Microbiol. **9:** 1193- 1201.

Fujiwara S., Fukuzawa H., Tachiki A., and Miyachi S. (1990) Structure and differential expression of two genes encoding carbonic anhydrase in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **87:** 9779-9783.

Fukuzawa H., Fujiwara S., Yamamoto Y., Dionisio-Sese M.L. and Miyachi S. (1990) cDNA cloning, sequence, and expression of carbonic anhydrase in *Chlamydomonas reinhardtii*-regulation by environmental $CO_2$ concentration. Proc. Natl. Acad. Sci. USA **87:** 4383-4387.

Fukuzawa H., Suzuki E., Komukai Y. and Miyachi S. (1992) A gene homologous to chloroplast carbonic anhydrase (*icfA*) is essential to photosynthetic carbon dioxide fixation by *Synechococcus* PCC 7942. Proc. Natl. Acad. Sci. USA **89:** 4437-4441.

Fukuzawa H., Miura K., Ishizaki K., Kucho K., Saito T., Kohinata T., Ohyama K. (2001) *Ccm1*, a regulatory gene controlling the induction of a carbon-concentrating mechanism in *Chlamydomonas reinhardtii* by sensing $CO_2$. Proc Natl. Acad. Sci. USA **98**: 5347-5352

Funke R.P., Kovar J.L. and Weeks D.P. (1997) Intracellular carbonic anhydrase is essential to photosynthesis in *Chlamydomonas reinhardtii* at atmospheric levels of $CO_2$. Plant Physiol. **114:** 237-244.

Geraghty A.M., Anderson J.C. and Spalding M.H. (1990) A 36 kilodalton limiting-$CO_2$ induced polypeptide of *Chlamydomonas* is distinct from the 37 kilodalton periplasmic carbonic anhydrase. Plant Physiol. **93:**116-121.

Gibbs S.P. (1962) The ultrastructure of the pyrenoids of green algae. J. Ultrastructure Res.**7:** 262-272.

Giordano M., Norici A., Forssen M., Eriksson M. and Raven J. (2003) An anaplerotic role for mitochondrial carbonic anhydrase in *Chlamydomonas reinhardtii*. Plant Physiol. **132:** 2126-2134.

Goodwin T.W. and Mercer E.I. (1983) Introduction to plant biochemistry. Pergamon press, New York, USA.

González-Meler M.A., Ribas-Carbo M., Siedow J.N. and Drake B.G. (1996) Direct inhibition of plant mitochondrial repiration by elevated $CO_2$. Plant Physiol. **112:** 1349-1355.

Gőtz R., Gnann A. and Zimmermann F.K. (1999) Deletion of the carbonic anhydrase-like gene NCE103 of the yeast *Saccharomyces cerevisiae* causes an oxygen-sensitive growth defect. Yeast, **15:** 855-864.

Goyal A. and Tolbert. N.E. (1988) Uptake of inorganic carbon by isolated chloroplasts from *Dunaliella tertiolecta*. Plant Physiol. **89:** 1264 -1269.

Guilloton M.B., Lamblin A.F., Kozliak E.I., Gerami-Nejad M., Tu C., Silverman D., Anderson P.M. and Fuchs J.A. (1993) A physiological role for cyanate-induced carbonic anhydrase in Escherichia coli. J. Bact. **175:** 1443-1451.

Hanson D.T., Andrews T.J. and Badger M.R. (2002) Variability of the $CO_2$ concentrating mechanism in hornworts (Anthocerotophyta). Funct. Plant Biol. **29:** 407-416.

Hanson D.T., Franklin L.A., Samuelsson G., Badger M.R. (2003) The *Chlamydomonas reinhardtii cia3* mutant lacking a thylakoid lumen-localized carbonic anhydrase is limited by $CO_2$ utilization by Rubisco and not PSII function *in vivo.* Plant Physiol. **132:** 2267-2275.

Hartman F.C. and Harpel M.R. (1994) Structure, function, regulation and assembly of D-ribulose-1,5-bisphosphate carboxylase/oxygenase. Ann. Rev. Biochem. **63:** 197-234.

Hatch M.D. (1987) $C_4$ photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim Biophys Acta. **895:** 81-106.

Henry R.P. (1996) Multiple roles of carbonic anhydrase in cellular transport and metabolism. Ann. Rev. Physiol. **58:** 523-538.

Hewett-Emmett D. and Tashian R.E. (1996) Functional diversity, conservation, and convergence in the evolution of the α-, ß-, and γ-carbonic anhydrase gene families. Mol. Phylogene. and Evol. **5:** 50-77.

Higgins C.F. (1992) ABC transporters: from microorganisms to man. Ann. Rev. Cell Biol. **8:** 67-113.

Hiltonen T., Björkbacka H., Forsman C., Clarke A.K. and Samuelsson G. (1998) Intracellular beta-carbonic anhydrase of the unicellular green alga *Coccomyxa*. Plant Physiol. **117:** 1341-1349.

Holden M. (1976) Chlorophylls. In: Chemistry and Biochemistry of Plant Pigments, Academic Press Inc., London, editors T.W.Goodwin. pp 2-37.

Honegger R. (1991) Functional aspects of the lichen symbiosis. Ann. Rev. Plant Physiol and Plant Mol. Biol. **42:** 553-578.

Ivanov S.V., Kuzmin I., Wei M.H., Pack S., Geil L., Johnson B.E., Stanbridge E.J. and Lerman M.I. (1998) Down-regulation of transmembrane carbonic anhydrases in renal cell carcinoma cell lines by wild-type von Hippel-Lindau transgenes. Proc. Natl. Acad. Sci. USA **95:** 12596-12601.

Jiang W.P. and Gupta D. (1999) Structure of the carbonic anhydrase VI (*CA6*) gene: evidence for two distinct groups within the alpha-CA gene family. Biochem. J. **344:** 385-390.

Johansson I.M. and Forsman C. (1993) Kinetic-studies of pea carbonic-anhydrase. Eur. J. Biochem. **218:** 439-446.

Jordan D.B. and Ogren W.L. (1981) Species variation in the specificity of ribulose bisphosphate carboxylase-oxygenase. Nature, **291:** 513-515.

Jordan D.B. and Ogren W.L. (1983) Species variation in the kinetic properties of ribulose 1, 5-bisphosphate carboxylase/oxygenase. Arch. Biochem. Biophys. **274:** 425- 433.

Kamo T., Shimogawara K., Fukuzawa H., Muto S. and Miyachi S. (1990) Subunit constitution of carbonic-anhydrase from *Chlamydomonas reinhardtii*. Eur. J. Biochem. **192:** 557-562.

Kaplan A. and Reinhold L. (1999) $CO_2$ concentrating mechanisms in photosynthetic microorganisms. Ann. Rev. Plant Physiol. and Plant Mol. Biol. **50:** 539-570.

Kaplan A., Schwarz R., Lieman-Hurwitz J. and Reinhold L. (1994) Physiological and molecular studies on the response of cyanobacteria to changes in the ambient inorganic carbon concentration. In: Molecular biology of the Cyanobacteria. editor, D. Bryant, Kluwer Academic Pub. Dordrecht, The Netherlands. pp. 469-485.

Karlsson J., Hiltonen T., Husic H.D., Ramazanov Z. and Samuelsson G. (1995) Intracellular carbonic anhydrase of *Chlamydomonas reinhardtii*. Plant Physiol. **109:** 533-539.

Karlsson J., Clarke A.K., Chen Z.Y., Hugghins S.Y., Park Y.I., Husic H.D., Moroney J.V. and Samuelsson G. (1998) A novel alpha-type carbonic anhydrase associated with the thylakoid membrane in *Chlamydomonas reinhardtii* is required for growth at ambient $CO_2$. EMBO J. **17:** 1208-1216.

Karhumaa P., Parkkila S., Waheed A., Parkkila A., Kaunisto K., Tucker P., Huang C., Sly W. and Rajaniemi H. (2000) Nuclear Non/p54 $^{nrb}$ protein is a nonclassical carbonic anhydrase. J. Biol. Chem. **275:** 16044-16049.

Katoh A., Lee K.S., Fukuzawa H., Ohyama K. and Ogawa T. (1996) *cemA* homologue essential to $CO_2$ transport in the cyanobacterium *Synechocystis* PCC 6803. Proc. Natl. Acad. Sci. USA **93:** 4006-4010.

Katzman G.L., Carlson S.J., Marcus Y., Moroney J.V. and Togasaki R.K. (1994) Carbonic anhydrase activity in isolated chloroplasts of wild-type and high-$CO_2$-dependent mutants of *Chlamydomonas reinhardtii* as studied by a new assay. Plant Physiol. **105:** 1197-1202.

Kavroulakis N., Flemetakis E., Aivalakis G. and Katinakis P. (2000) Carbon metabolism in developing soybean root nodules: The role of carbonic anhydrase. Mol. Plant-Microbe Interact. **13:** 14-22.

Keys A.J. (1986) Rubisco: its role in photorespiration. Phil. Trans. R. Soc. Lon. Ser B. **313:** 325-336.

Khalifah, R.G. (1971) The carbon dioxide hydration activity of carbonic anhydrase, Stop-flow kinetic studies on the native human isoenzymes B and C. J. Biol. Chem. **246:** 2561-2573.

Kim H.J. (1997) Identification of the physiological role of carbonic anhydrase using the antisense technique and investigation of its transcriptional regulation in *Arabidopsis thaliana* (L.) heynh. Louisiana State University, Dissertation.

Kimber M.S., and Pai E.F. (2000) The active site architecture of *Pisum sativum* β-carbonic anhydrase is a mirror image of that of α-carbonic anhydrases. EMBO J. **19:** 1407-1418.

Kimpel D.L., Togasaki R.K. and Miyachi S. (1983) Carbonic anhydrase in *Chlamydomonas reinhardtii*. I: Localization. Plant Cell Physiol. **24:** 255-259.

Kisker C., Schindelin H., Alber B.E., Ferry J.G. and Rees D.C. (1996) A left-handed beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila,* EMBO J. **15:** 2323-2330.

Kuchitsu K. Tsuzuki M. and Miyachi S. (1991) Polypeptide composition of the pyrenoid and its regulation by $CO_2$ concentration in unicellular green algae. Can. J. Bot. **69:** 1062-1069.

Kulp D., Haussler D., Reese M.G., and Eeckman F. (1996) In: ISMB-96, AAAI Press, pp. 134-142, http://www.cse.ucsc.edu/~dkulp/cgi-bin/greenGenie.

Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, **227:** 680-685.

Lane T.W. and Morel F.M.M. (2000) A biological function for cadmium in marine diatoms. Proc Natl. Acad. Sci. USA. **97:** 4627-4631.

Leggat W., Badger M.R. and Yellowlees D. (1999) Evidence for an inorganic carbon-concentrating mechanism in the symbiotic dinoflagellate *Symbiodinium*. Plant Physiol. **121 (4):** 1247-1255.

Lindskog S. (1997) Structure and mechanism of carbonic anhydrase. Pharmacol. Ther. **74:** 1-20.

Lowry O.H, Rosebrough N.J., Farr A.L. and Randall R.J. (1951) Protein measurement with Folin phenol reagent. J. Biol.Chem. **193:** 265-275.

Ludwig M. and Burnell J.N. (1995) Molecular comparison of carbonic anhydrase from *Flaveria* species demonstrating different photosynthetic pathways. Plant Mol. Biol. **29:** 353-365.

Maeda S., Badger M.R. and Price G.D. (2001) Novel gene products associated with NdhD3/D4 containing NDH-I complexes are involved in photosynthetic $CO_2$ hydration in the cyanobacterium, *Synechococcus sp*. PCC7942. Mol. Microbiol. **43:** 425-435.

Majeau N., Arnoldo M. and Coleman J.R. (1994) Modification of carbonic anhydrase activity by antisense and over-expression constructs in transgenic tobacco. Plant Mol. Biol. **25:** 377-385.

Manuel L.J. and Moroney J.V. (1988) Inorganic carbon accumulation by *Chlamydomonas reinhardtii*; new proteins are made during adaptation to low $CO_2$. Plant Physiol. **88:** 491-496.

Mason C.B., Manuel L.J. and Moroney J.V. (1990) A new chloroplast protein is induced by growth on low $CO_2$ in *Chlamydomonas reinhardtii*. Plant Physiol. **93:** 833-836.

McKay R.M.L. and Gibbs S.P. (1991) Composition and function of pyrenoids: cytochemical and immunocytochemical approaches. Can. J. Bot. **69:** 1040-1052.

McKay R.M.L., Gibbs S.P. and Espie G.S. (1993) Effect of dissolved inorganic carbon on the expression of carboxysomes, localization of rubisco and the mode of inorganic carbon transport in cells in the cyanobacterium *Synechococcus* UTEX 625. Arch. Microbiol. **159:** 21-29.

Mckay R.M.L., Gibbs S.P. and Vaughan K.C. (1991) Rubisco activase is present in the green algae. Protoplasma, **162:** 38-45.

Meldrum N.U. and Roughton, F.J.W. (1933) Carbonic anhydrase. Its preparation and properties. J. Physiol. **80:** 113-142.

Mitsuhashi S. and Miyachi S. (1996) Amino acid sequence homology between N- and C-terminal halves of a carbonic anhydrase in *Porphyridium purpureum*, as deduced from the cloned cDNA. J. Biol. Chem. **271:** 28703-28709.

Mitsuhashi S., Mizushima T., Yamashita E., Yamamoto M., Kumasaka T., Moriyama H., Ueki T., Miyachi S. and Tsukihara T. (2000) X-ray structure of beta-carbonic anhydrase from the red alga, *Porphyridium purpureum*, reveals a novel catalytic site for $CO_2$ hydration. J. Biol. Chem. **275:** 5521-5526.

Mitsuhashi S., Ohnishi J., Hayashi M. and Ikeda M. (2003) A gene homologous to beta-type carbonic anhydrase is essential for the growth of *Corynebacterium glutamicum* under atmospheric conditions. Appl. Microbiol. Biotechnol. (in press).

Morel F.M.M., Cox E.H., Kraepiel A.M.L., Lane T.W., Milligan A.J., Schaperdoth I., Reinfelder J.R. and Tortell P.D. (2002) Acquisition of inorganic carbon by the marine diatom *Thalassiosira weissflogii*. Funct. Plant Biol. **29:** 301-308.

Morita E., Kuroiwa H., Kuroiwa T. and Nozaki H. (1997) High localization of ribulose 1,5-bisphosphate carboxylase/oxygenase in the pyrenoids of *Chlamydomonas reinhardtii* (Chlorophyta) as revealed by cryofixation and immunogold electron microscopy. J. Phycol. **33:** 68-72.

Moroney J.V. and Mason C.B. (1991) The role of chloroplast in inorganic carbon acquisition by *Chlamydomonas reinhardtii*. Can. J. Bot. **69:** 1017-1024.

Moroney J.V. and Somanchi A. (1999) How do algae concentrate $CO_2$ to increase the efficiency of photosynthetic carbon fixation? Plant Physiol. **119:** 9-16.

Moroney J.V. and Tolbert N.E. (1985) Inorganic carbon uptake by *Chlamydomonas reinhardtii*. Plant Physiol. **77:** 253-258.

Moroney J.V., Bartlett S.G. and Samuelsson G. (2001) Carbonic anhydrases in plants and algae. Plant Cell Environ. **24:** 141-153.

Moroney J.V., Husic H.D. and Tolbert N.E. (1985) Effect of carbonic anhydrase inhibitors on inorganic carbon accumulation by *Chlamydomonas reinhardtii*. Plant Physiol. **79:** 177-183.

Moroney J.V., Husic H.D. and Tolbert N.E. (1986) $CO_2$ and $HCO_3^-$ accumulation by microalgae. In : Regulation of chloroplast differentiation, editors, Akoyunoglou G., Seneger H., Alan R. Liss, New York. pp 715-724.

Moroney J.V., Tolbert N.E. and Sears B.B. (1986) Complementation analyses of the inorganic carbon concentrating mechanism of *Chlamydomonas reinhardtii*. Mol. General Gen. **204:** 199-203.

Moroney J.V., Togasaki R.K., Husic H.D. and Tolbert N.E. (1987) Evidence that an internal carbonic anhydrase is present in 5% $CO_2$ grown and air grown *Chlamydomonas*. Plant Physiol. **84:** 757-761.

Newman S.M., Boynton J.E., Gillham N.W., Randolph-Anderson B.L., Johnson A.M. and Harris E.H. (1990) Transformation of chloroplast ribosomal-RNA genes in *Chlamydomonas*-molecular and genetic-characterization of integration events. Genetics, **126:** 875-888.

Newman T., DeBruijn F.J., Green P., Keegstra K., Kende H., McIntosh L., Ohlrogge J., Raikhel N., Somerville S., Thomashow M., Retzel E. and Somerville C. (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol. **106:** 1241-1255.

Nickrent D.L., Parkinson C.L., Palmer J.D. and Duff R.J. (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol. Biol. Evol. **17:** 177-183.

Niles E.G., Condit R.C., Caro P., Davidson K., Matusick L. and Seto J. (1986) Nucleotide sequence and genetic map of the 16-kb vaccinia virus HindIII D fragment. Virology, **153:** 96-112.

Okada M. (1992) Recent studies on the composition and activity of algal pyrenoids. In: Progress in Phycological Research, Vol 8, Editors, Round F.E. and Chapman D.J. Biopress, Ltd. Bristol, UK. pp. 117-138.

Okamura M., Price G.D., Badger M.R., Ogawa T. and Omata T. (1997) The *cmpABCD* genes of the cyanobacterium *Synechococcus* PCC 7942 encode a $HCO_3^-$ transporter. Plant Cell Physiol. **38:** supplement, 30.

Omata T. (1991) Cloning and characterization of the *nrtA* gene that encodes a 45 kDa protein involved in nitrate transport in the cyanobacterium *Synechococcus* PCC 7942. Plant Cell Physiol. **32:** 151-157.

Omata T., Price G.D., Badger M.R., Okamura M., Gohta S. and Ogawa T. (1999) Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp strain PCC 7942. Proc. Natl. Acad. Sci., USA **96:** 13571-13576.

Orús M.I., Rodríguez M.L., Martínez F., Marco E. (1995) Biogenesis and ultrastructure of carboxysomes from wild-type and mutants of *Synechococcus* sp strain PCC-7942. Plant Physiol. **107:** 1159-1166.

Osterlind S. (1950) Inorganic carbon sources of green algae. I: Growth experiments with *Scenedesmus quadricauda* and *Chlorella pyrenoidosa*. Physiol. Plant. **3:** 353-360.

Palmqvist K., Yu J.W. and Badger M.R. (1994) Carbonic anhydrase activity and inorganic carbon fluxes in low and high $C_i$ cells of *Chlamydomonas reinhardtii* and *Scenedesmus obliquus*. Physiol. Plant. **90:** 537-547.

Park Y.I., Karlsson J., Rojdestvenski I., Pronina N., Klimov V., Öquist G. and Samuelsson G. (1999) Role of a novel photosystem II-associated carbonic anhydrase in photosynthetic carbon assimilation in *Chlamydomonas reinhardtii*. FEBS Lett. **444:** 102-105.

Price G.D. and Badger M.R. (1989a) Expression of human carbonic-anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high $CO_2$-requiring phenotype-evidence for a central role for carboxysomes in the $CO_2$ concentrating mechanism. Plant Physiol. **91,** 505-513.

Price G.D. and Badger M.R. (1989b) Isolation and characterization of high-$CO_2$ requiring mutants of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate $CO_2$ within the carboxysome. Plant Physiol. **91:** 514-525.

Price G.D., Coleman J.R. and Badger M.R. (1992) Association of carbonic-anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. Plant Physiol. **100:** 784-793.

Price G.D., Howitt S.M., Harrison K. and Badger M.R. (1993) Analysis of a genomic DNA region from the cyanobacterium *Synechococcus* sp strain PCC7942 involved in carboxysome assembly and function. J. Bact. **175:** 2871-2879.

Price G.D., Sültemeyer D., Klughammer B., Ludwig M. and Badger M.R. (1998) The functioning of the $CO_2$ concentrating mechanism in several cyanobacterial strain: a review of general physiological characteristics, genes, proteins, and recent advances. Can. J. Bot. **76:** 973-1002.

Price G.D., Maeda S., Omata T., and Badger M.R. (2003) Modes of active inorganic carbon uptake in the cyanobacterium Synechococcus PCC7942 (in press).

Price G.D., von Caemmerer S., Evans J.R., Yu J.W., Lloyd J., Oja V., Kell P., Harrison K., Gallagher A. and Badger M.R. (1994) Specific reduction of chloroplast carbonic anhydrase activity by antisense RNA in transgenic tobacco plants has a minor effect on photosynthetic $CO_2$ assimilation. Planta, **193:** 331-340.

Purton S. and Rochaix J.P. (1994) Complementation of a *Chlamydomonas reinhardtii* mutant using a genomic cosmid library. Plant Mol Biol. **24:** 533-537.

Ramazanov Z., Mason C.B., Geraghty A.M., Spalding M.H. and Moroney J.V. (1993) The low $CO_2$ inducible 36-kilodalton protein is localized to the chloroplast envelope of *Chlamydomonas reinhardtii.* Plant Physiol. **101:** 1195-1199.

Ramazanov Z., Rawat M., Henk M.C., Mason C.B. and Moroney J.V. (1996) Ultrastructural and biochemical adaptation of algal cells to limiting $CO_2$ concentrations. Scientia Marina (Suppl.1), **60:** 141-148.

Raven J.A. (1997) $CO_2$-concentrating mechanisms: A direct role for thylakoid lumen acidification? Plant Cell Environ. **20:** 147-154.

Raven J.A. and Newman J.R. (1994) Requirement for carbonic-anhydrase activity in processes other than photosynthetic inorganic carbon assimilation. Plant Cell Environ. **17:** 123-130.

Raven J.A. (2001) A role for mitochondrial carbonic anhydrase in limiting $CO_2$ leakage from low $CO_2$-grown cells of *Chlamydomonas reinhardtii*. Plant Cell Environ. **24:** 261-265.

Rawat M. and Moroney J.V. (1991) Partial characterization of a new isoenzyme of carbonic-anhydrase isolated from *Chlamydomonas reinhardtii*. J. Biol. Chem. **266:** 9719-9723.

Rawat M., Henk M.C., Lavigne L.L. and Moroney J.V. (1996) *Chlamydomonas reinhardtii* mutants without ribulose-1,5-bisphosphate carboxylase-oxygenase lack a detectable pyrenoid. Planta, **198:** 263-270.

Reinfelder J.R., Kraepiel A.M.L. and Morel F.M.M. (2000) Unicellular $C_4$ photosynthesis in a marine diatom. Nature, **407:** 996-999.

Reinhold L., Hosloff R. and Kaplan A. (1991) A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. Can. J. Bot. **69:** 984-988.

Roberts S.B., Lane T.W. and Morel F.M.M. (1997) Carbonic anhydrase in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae). J. Phycol. **33:** 845-850.

Ronen-Tarazi M., Bonfil D.J., Schatz D., Kaplan A. (1998) Cyanobacterial mutants impaired in bicarbonate uptake isolated with the aid of an inactivation library. Can.J. Bot. **76:** 942-948.

Ronen-Tarazi M., Lieman–Hurwitz J., Gabay C., Orús M.I. and Kaplan A. (1995) The genomic region of rbcLS in *Synechococcus* PCC 7942 contains genes involved in the ability to grow under low $CO_2$ concentration and in chlorophyll biosynthesis. Plant Physiol. **108:** 1461-1469.

Rowlett R.S., Chance M.R., Wirt M.D., Sidelinger D.E., Royal J.R., Woodroffe M., Wang Y.F.A., Saha R.P. and Lam M.G. (1994) Kinetic and structural characterization of spinach carbonic-anhydrase. Biochem. **33:** 13967-13976.

Sambrook J., Fritsch E.F. and Maniatis T. (1989) 'Molecular Cloning: A Laboratory Manual, ($2^{nd}$ Edn)' (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY)

Satoh D., Hiraoka Y., Colman B. and Matsuda Y. (2001) Physiological and molecular biological characterization of intracellular carbonic anhydrase from the marine diatom *Phaeodactylum tricornutum*. Plant Physiol. **126:** 1459-1470.

Schneider G., Lindqvist Y., Branden C.I. (1992) Rubisco-structure and mechanism. Ann. Rev. Biophys. Biomol. Struct. **21:** 119-143 1992.

Schnell R.A. and Lefebvre P.A. (1993) Isolation of the *Chlamydomonas reinhardtii* regulatory gene NIT2 by transposon tagging. Genetics, **134**: 737-747

Schwartz R., Reinhold L. and Kaplan A. (1995) Low activation state of 1,5-bisphosphate carboxylase/oxygenase activation in carboxysome defective *Synechococcus* mutants. Plant Physiol. **108:** 183-190.

Shibata M., Ohkawa H., Kaneko T., Fukuzawa H., Tabata S., Kaplan A. and Ogawa T. (2001) Distinct constitutive and low $CO_2$ induced $CO_2$ uptake systems in cyanobacteria: Genes involved and their phylogenetic relationship with homologous genes in other organisms. Proc Natl. Acad. Sci. USA **98:** 11789-11794.

Shibata M., Ohkawa H., Shimoyama M. and Ogawa T. (2002) Two $CO_2$ uptake systems: four systems for inorganic carbon acquisition in *Synechocystis sp*. Strain PCC 6803. Funct. Plant Biol. **29:** 123- 129.

Silflow C.D. (1998) Organization of the Nuclear Genome. In: The Molecular Biology of Chloroplasts and Mitochondria in *Chlamydomonas*, editors, Rochaix J.D., Goldschmidt-Clermont M. and Merchant S., Kluwer Academic Publishers.

Slaymaker D.H., Navarre D.A., Clark D., Pozo O., Martin G.B. and Klessig D. (2001) The tobacco salicylic acid binding protein 3 (SABP3) is the chloroplast carbonic anhydrase which exhibits antioxidant activity and plays a role in the hypersensitive defense response. Proc Natl. Acad. Sci. USA **99:** 11640-11645.

Sly W.S. and Hu P.Y. (1995) Human carbonic anhydrases and carbonic anhydrase deficiencies. Ann. Rev. Biochem. **64:** 375-401.

Smith K.S., and Ferry J.G. (1999) A plant-type (beta-class) carbonic anhydrase in the thermophilic methanoarchaeon *Methanobacterium thermoautotrophicum*. J. Bact. **181:** 6247-6253.

Smith E.C. and Griffiths H. (1996) A pyrenoid based carbon concentrating mechanism is present in terrestrial bryophytes of the class anthocerotae. Planta, **200:** 203-212.

Smith E.C. and Griffiths H. (2000) The role of carbonic anhydrase in photosynthesis and the activity of the carbon concentrating mechanism in bryophytes of the class Anthocerotae. New Phytol. **145:** 29-37.

So A.K.C., Cot S.S.W and Espie G.S. (2002) Characterization of the C-terminal extension of carboxysomal carbonic anhydrase from *Synechocystis* sp. PCC6803. Funct. Plant Biol. **29:** 183-194.

Soltes-Rak E., Mulligan M.E. and Coleman J.R. (1997) Identification and characterization of a gene encoding a vertebrate-type carbonic anhydrase in cyanobacteria. J. Bact. **179:** 769-774.

Spalding M.H. and Jeffrey M. (1989) Membrane-associated polypeptides induced in *Chlamydomonas* by limiting $CO_2$ concentrations. Plant Physiol. **89:**133-137.

Spalding M.H., Spreitzer R.J., Ogren W.L. (1983) Carbonic anhydrase deficient mutant of *Chlamydomonas* requires elevated carbon dioxide concentration for photoautotrophic growth. Plant Physiol. **73:** 268-272

Strayer D.S. and Jerng H.H. (1992) Sequence and analysis of the BamH1 "D" fragment of Shope fibroma virus: Comparison with similar regions of related poxviruses. Virus Res. **25:** 117-132.

Sueoka, N. (1960). Mitotic replication of deoxyribonucleic acids in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **46:** 83-91.

Sültemeyer D.F., Miller A.G., Espie G.S., Fock H.P. and Canvin D.T. (1989) Active $CO_2$ transport by the green algae *Chlamydomonas reinhardtii*. Plant Physiol. **94:** 1250- 1257.

Sültemeyer D.F., Klughammer B., Ludwig M., Badger M.R. and Price G.D. (1997) Random insertional mutagenesis used in the generation of mutants of the marine cyanobacterium *Synechococcus* sp. Strain PCC 7002 with an impaired $CO_2$ concentrating mechanism. Aust. J. Plant Physiol. **24:** 317-327.

Sültemeyer D.F., Price G.D, Yu J.W., and Badger M.R. (1995) Characterization of carbon dioxide and bicarbonate transport during steady state photosynthesis in the marine cyanobacterium *Synechococcus strain* PCC 7002. Planta, **197:** 597-607.

Thielmann J., Tolbert N.E, Goyal A. and Senger H. (1990) Two systems for concentrating $CO_2$ and bicarbonate during photosynthesis by *Scenedesmus*. Plant Physiol. **92:** 622-629.

Tobin A.J. (1970) Carbonic anhydrase from parsley leaves. J. Biol.Chem. **245:** 2656- 2666.

Tsuzuki M. (1983) Mode of $HCO_3^-$-utilization by the cells of *Chlamydomonas reinhardtii* grown under ordinary air. Z. Pflanzenphysiol. **110:** 29-37

Türeci O., Sahin U., Vollmar E., Siemer S., Gottert E., Seitz G., Parkkila A.K., Shah G.N., Grubb J.H., Pfreundschuh M., Sly W.S. (1998) Human carbonic anhydrase XII: cDNA cloning, expression, and chromosomal localization of a carbonic anhydrase gene that is overexpressed in some renal cell cancers Proc. Natl. Acad. Sci. USA **95:** 7608-7613.

Turpin D.H., Miller A.G., and Canvin D.T. (1984) Carboxysome content of *Synechococcus leopoliensis* (Cyanophyta) in response to inorganic carbon. J. Phycol. **20:** 249-253.

Van Hunnik E., Livine A., Pogenberg V., Spijkerman E., van den Ende H., Mendoza E.G., Sültemeyer D. and de Leeuw J.W. (2001) Identification and localization of a thylakoid bound carbonic anhydrase from the green alga *Tetraedron minimum* (Chlorophyta) and *Chlamydomonas noctigama* (Chlorophyta). Planta, **212:** 454-459.

Van K. and Spalding M.H. (1999) Periplasmic carbonic anhydrase structural gene (*Cah1*) mutant in *Chlamydomonas reinhardtii*. Plant Physiol. **120:** 757-764.

Villarejo A., Rolland N., Martinez F. and Sültemeyer D. (2001) A new chloroplast envelope carbonic anhydrase activity is induced during acclimation to low inorganic carbon concentrations in *Chlamydomonas reinhardtii*. Planta, **213:** 286-295.

Villarejo A., Shutova T., Moskvin O. Forssen M., Klimov V.V. and Samuelsson G. (2002) A photosystem II-associated carbonic anhydrase regulates the efficiency of photosynthetic oxygen evolution. EMBO J. **21:** 1930-1938.

Wilbur K.M. and Anderson N.G. (1948) Electrometric and colorometric determination of carbonic anhydrase. J. Biol. Chem. **176:** 147-154.

Xiang Y.B., Zhang J. and Weeks D.P. (2001) The *Cia5* gene controls formation of the carbon concentrating mechanism in *Chlamydomonas reinhardtii*. Proc. Natl. Acad. Sci. USA **98:** 5341-5346.

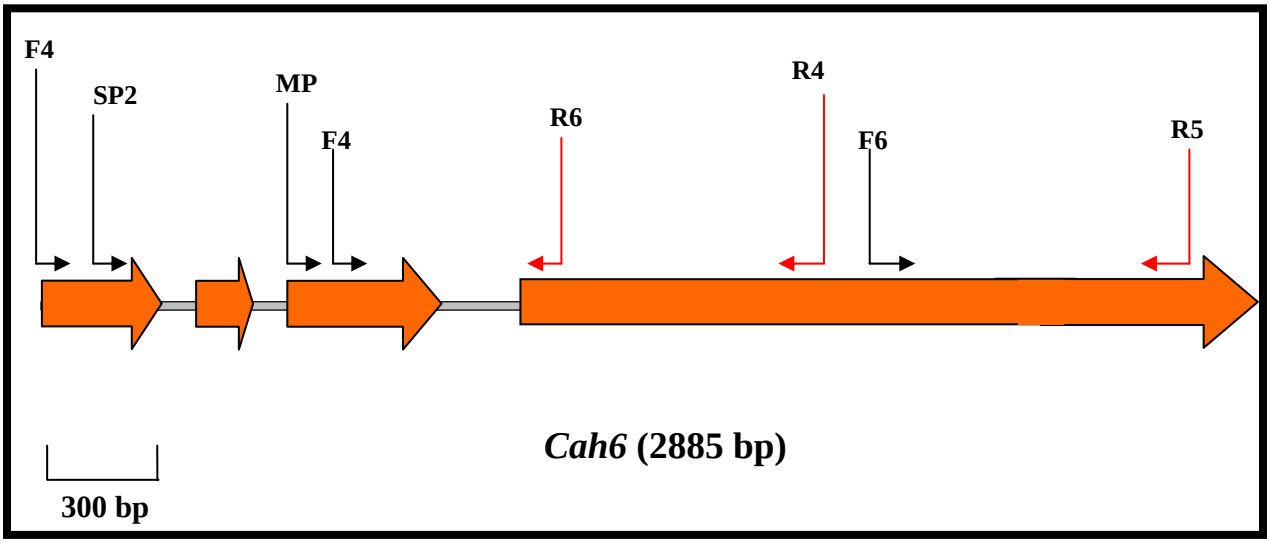
Yang Z., Zhang Q. and Xu Z. (1999) Cloning of a cDNA of a duplicated carbonic anhydrase in *Dunaliella salina.* (Accession number AAF22644).

Yu J.W, Price G.D., Song L. and Badger M.R. (1992) Isolation of a putative carboxysomal carbonic anhydrase gene from the cyanobacterium *Synechococcus* PCC7942. Plant Physiol. **100:**794-800.

Yu J.W., Price G.D. and Badger M.R. (1994) Characterization of $CO_2$ and $HCO_3^-$ uptake during steady state photosynthesis in the cyanobacterium *Synechococcus* PCC 7942. Aust. J. Plant Physiol. **21:** 185-195.

# APPENDIX 1

## *Cah6* MAP AND PRIMERS




The orange block arrows and interconnecting lines represent the exons and the introns, respectively. The 5′ end and the 3′ end primers are denoted by black and red arrows, respectively.

5′ end primers:- (5′ → 3′)

| | |
|---|---|
| MP | AGCAACCGCAGCAGCCTT |
| SP2 | ATGGGATGCGGTGCCAGCGTG |
| F4 | GCACGAGGCAACATTAAACA |
| F5 | GCATCCCCACCAGGACTTCA |
| F6 | CCTGAGCAGGGTGAATGGAA |
| X9 | AAACTCAACTCCTTCATAATAGGC |

3′ end primers:- (5′ → 3′)

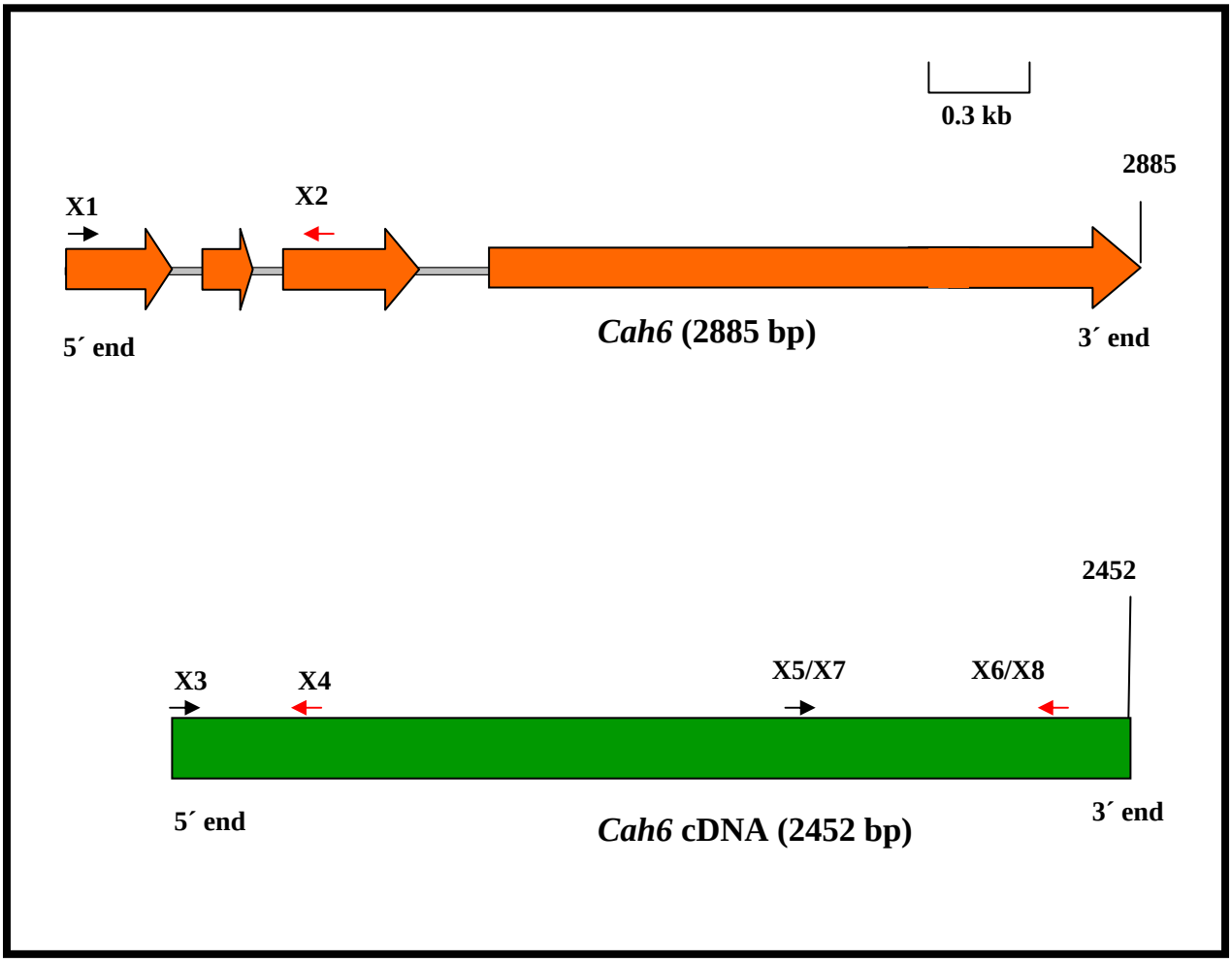
| | |
|---|---|
| R4 | TTGCGCCATGAAGTCCCTAA |
| R6 | TTGACGTTCTCCTTCACCAC |
| R5 | TGCGGTACAGATTACAGTCA |

R4H ATATAAGCTTTTGCGCCATGAAGTCCCTAA

The primer R4H was generated by adding a HindIII site (shown in blue) and four extra bases (shown in black) at the 5′ end of the primer R4.

# APPENDIX 2

## *Cah6* MAP AND RNAi PRIMERS




The orange block arrows and interconnecting lines represent the exons and the introns, respectively. The 5´ end and the 3´ end primers are denoted by black and red arrows, respectively.

5′ end primers: - (5′ → 3′)

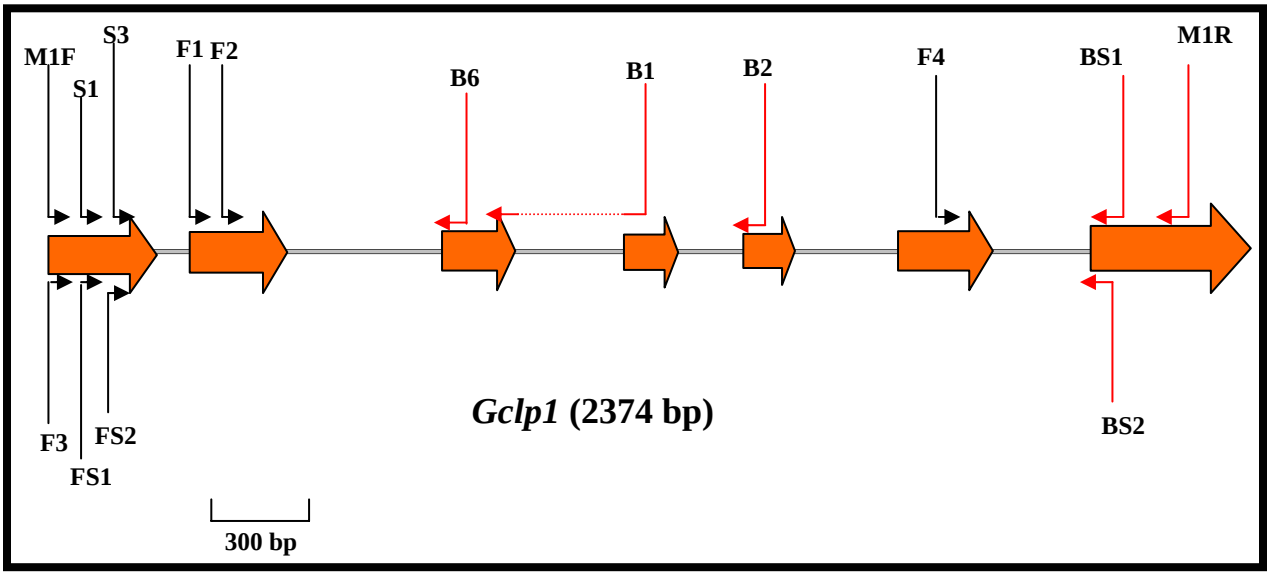
| | |
|---|---|
| X1 | TCCCCCGGGCAGATGATACCAGCAACATTAAACA |
| X3 | CGGGATCCCGTGAGCGAATTGATCTTTTCA |
| X5 | TCCCCCGGGAAACTCAACTCCTTCATAATAGGC |
| X7 | CGGGATCCCCTCGCAGATACATCCCT |

3′ end primers: - (5′ → 3′)

| | |
|---|---|
| X2 | CGGAATTCCGTGAGCGAATTGATCTTTTCA |
| X4 | GCTCTAGACAGATGATACCAGCAACATTAAACA |
| X6 | CGGAATTCCCTCGC AGA TAC ATC CCT |
| X8 | GCTCTAGAAAACTCAACTCCTTCATAATAGGC |

# APPENDIX 3

## *Gclp1* MAP AND PRIMERS




The orange block arrows and lines interconnecting the block arrows represent the exons and introns, respectively. 5′ end primers and 3′ end primers are shown by black and red arrows, respectively. The broken line in B1 primer denotes the missing DNA sequence in the primer.

5′ end primers: - (5′ → 3′)

| | |
|---|---|
| M1F | GCAGTTAGTCCATGCGAA |
| F1 | CCGCAAGCCGGATGTGCCCGTTAA |
| F2 | GGGGGCGGGCTCATCGGTGTGGTATG |
| FS1 | GCTTTCTACGTGACACAAGG |
| FS2 | GGACCATGTCGCTATTCA |
| F4 | GCTGCTACGCCGAGCTGTC |
| S1 | ATGTCGCTATTCAAGTCTAGCCTG |
| S3 | GCGCTGGATGAGCTGGGC |
| F3 | CCTTGTGCGGGGCTTTCTAC |

3′ end primers: - (5′ → 3′)

| | |
|---|---|
| M1R | TTGAATGAGTACGGTGCA |
| BS1 | GTAAAGAGAGGAGCTGCAAG |
| BS2 | GCCAAGAGGTAAAGAGAGGA |
| B1 | CAATGTGACCATTGGCCACGCCGC |
| B2 | ACACCACCATCCCCTCGGGCCAG |
| B6 | TGTCCAAGTACAGCATGG |
| M1RH | ATATAAGCTT TTGAATGAGTACGGTGCA |

FS1rev, FS2 rev, F1 rev, F2rev, B4, B6rev, M1F rev, M1R rev, B1rev, B2rev, B4 are reverse complementary primers of FS1, FS2, F1, F2, F4, B6, M1F, M1R, B1, B2 and F4, respectively. The primer R4H was generated by adding a HindIII site (shown in blue) and four extra bases (shown in black) at the 5′ end of the primer R4.

# VITA

Mautusi Mitra was born in Calcutta, India, on February 14, 1968**.** She attended Presidency College at Calcutta, India, majoring in botany. She went on to complete her Bachelor of Science degree in botany at the University of Calcutta, India, in 1989. She pursued a Master of Science degree in botany at the University of Calcutta, India, which she received in 1991. After receiving her master's degree, she pursued personnel management for five years. In 1996, Mautusi received a Junior Research Fellowship from the Council of Scientific and Industrial research in India. She then attended the Graduate School at Louisiana State University to pursue a doctoral degree in plant biology. Her dissertation research has dealt with the identification, cloning and overexpression of carbonic anhydrase and carbonic anhydrase like genes of the green alga *Chlamydomonas reinhardtii*, under the supervision of Dr. J.V. Moroney. After her graduation, Mautusi, is going to join Dr. Anastasios Melis's laboratory at the University of California, Berkeley, where she will be doing research on photosystem II repair mechanisms and hydrogen production in green algae.